

SEARCH REQUEST FORM

Requestor's Name: D. B. Jahan Serial Number: 08/483,534
Date: 04/16/98 Phone: 305-2150 Art Unit: 1646

Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors, keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

Please search

AA 1-168 SEQ ID NO: 2 NA 94-597 SEQ ID NO: 1
" 1-307 SEQ ID NO: 2 " 1-597 SEQ ID NO: 1

only goes to 168
Thanks.

STAFF USE ONLY

Date completed: 4-28-98
Searcher: MANIC
Terminal time: 15
Elapsed time: 15
CPU time: 5
Total time: 20
Number of Searches: 1
Number of Databases: 7

Search Site

☒ STIC
☒ CM-1
☐ Pre-S

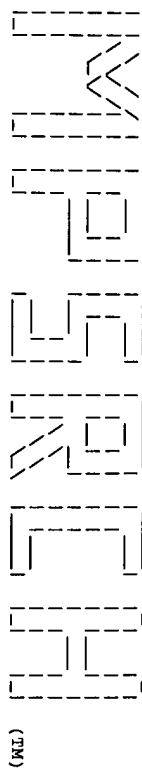
Type of Search

☒ N.A. Sequence
☒ A.A. Sequence
☐ Structure
☐ Bibliographic

Vendors

☒ STIC
☒ STN
☒ Dialog
☐ APS
☐ Geninfo
☐ SDC
☐ DARC/Questel
☐ Other

THIS PAGE BLANK (USPTO)



Release 3.0.4A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1997 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch.p protein - protein database search, using Smith-Waterman algorithm
Run on: Sat Apr 18 02:02:33 1998; Maspar time 10.96 Seconds
Tabular output not generated. 645.537 Million cell updates/sec

Title: >US-08-483-534A-2
Description: (1-168) from US08483534A.pep
Perfect Score: 1198
Sequence: 1 EEVIPSRLDIRVGKITVEK.....NEMTKLSISCKSLKGNIS 168

Scoring table: PAM 150
Gap 11

Searched: 140555 seqs, 42109429 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl5
1:sp.fungi 2:sp.human 3:sp.invertebrate 4:sp.mammal
5:sp.mhc 6:sp.organelle 7:sp.phage 8:sp.plant
9:sp.bacteria 10:sp-rodent 11:sp.virus 12:sp-vertebrate
13:sp.unclassified

Statistics: Mean 42.115; Variance 76.839; scale 0.548

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	596	49.7	310	10	Q60659	4.63e-103
2	595	49.7	310	10	Q60659	4.63e-103
3	588	49.1	312	2	Q12904	2.76e-101
4	263	22.0	658	9	Q28819	1.34e-31
5	236	19.7	629	9	Q33925	3.34e-26
6	222	18.5	651	9	Q26687	1.88e-23
7	138	11.5	246	9	Q27056	5.90e-08
8	108	9.0	201	9	Q34943	4.08e-03
9	100	8.3	336	9	Q28306	6.26e-02
10	99	8.3	395	9	Q52942	8.72e-02
11	99	8.3	832	9	P74619	8.72e-02
12	98	8.2	707	3	Q24428	1.21e-01
13	97	8.1	498	8	Q41597	1.69e-01
14	97	8.1	804	9	P94540	1.69e-01
15	94	7.8	1036	9	Q29316	4.45e-01
16	93	7.8	2408	2	Q92566	6.13e-01
17	94	7.8	2442	2	Q92793	4.45e-01
18	92	7.7	560	3	Q21237	8.41e-01
19	92	7.7	575	3	P90900	8.41e-01
20	92	7.7	611	8	Q41042	8.41e-01

21	91	7.6	120	9	Q31353	BC48A PROTEIN (FRAGMEN	1.15e+00
22	91	7.6	273	3	Q19694	SIMILAR TO C. ELEGANS	1.15e+00
23	91	7.6	1625	8	Q08367	ACETYL-COA CARBOXYLASE	1.15e+00
24	91	7.6	2325	8	Q41743	ACETYL-COENZYME A CARB	1.15e+00
25	90	7.5	358	4	Q28619	PROTEIN CO-FACTOR.	1.57e+00
26	90	7.5	462	9	Q05433	ATP SYNTHASE SUBUNIT B	1.57e+00
27	90	7.5	542	3	Q25618	INTERMEDIATE FILAMENT	1.57e+00
28	90	7.5	613	3	Q27334	OVIF.	1.57e+00
29	90	7.5	635	8	Q40363	NIM1 PROTEIN.	1.57e+00
30	90	7.5	1744	3	Q17008	T23B12.9 PROTEIN.	1.57e+00
31	90	7.5	1879	3	Q19161	COSMID F07C7.	1.57e+00
32	89	7.4	152	9	P75893	FROM BASES 1061668 TO	2.14e+00
33	89	7.4	179	3	Q22285	T07C12.7.	2.14e+00
34	89	7.4	418	9	Q06142	HYDROTHERMAL 44.9 KD P	2.14e+00
35	89	7.4	427	4	Q28349	VON WILLEBRAND FACTOR	2.14e+00
36	89	7.4	443	9	Q30820	ADENYLIC CYCLASE.	2.14e+00
37	89	7.4	610	3	Q17047	CYTOPLASMIC INTERMEDIA	2.14e+00
38	87	7.3	125	11	Q88488	NONSTRUCTURAL PROTEIN	3.95e+00
39	87	7.3	162	10	P70273	PROTEIN KINASE, MITOGE	3.95e+00
40	87	7.3	237	10	Q60521	PROTEIN KINASE, MITOGE	3.95e+00
41	87	7.3	334	10	P70236	PROTEIN KINASE, MITOGE	3.95e+00
42	87	7.3	498	6	Q19131	ADENOSINE TRIPHOSPHATA	3.95e+00
43	87	7.3	587	9	Q55312	SIMILAR TO E. COLI GLU	3.95e+00
44	87	7.3	595	11	Q84421	HYDROTHERMAL 292.7 KD	2.91e+00
45	88	7.3	2756	9	Q33904		

ALIGNMENTS

RESULT 1
ID Q60659 PRELIMINARY; PRT; 310 AA.

AC 060659;
DT 01-NOV-1996 (TREMBLER. 01, CREATED)
DT 01-NOV-1996 (TREMBLER. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLER. 05, LAST ANNOTATION UPDATE)
DE ENDOTHELIAL MONOCYTE ACTIVATING POLYPEPTIDE 2 (ENDOTHELIAL-MONOCYTE
DE ACTIVATING POLYPEPTIDE II).
GN ENAP2.
OS MUS MUSCULUS (MOUSE).
CC EUKARIOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
CC EUTHERIA; RODENTIA.
OC [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 95014290.
RA KAO J., HOUCK K., FAN Y., HAEHNEL I., LIBRETTI S.K., KAYTON M.L.,
RA GRKSCHETT T., CHABOT J., NORTGROD R., GREENBERG S., KUONG W.J.,
RA LEUNG D.W., HAYWARD J.R., KISIEL W., HEATH M., BRETT J.,
RA STERN D.M.;
RL J. BIOL. CHEM. 269:25106-25119(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94193665.
RA KAO J., FAN Y., HAEHNEL I., BRETT J., GREENBERG S., CLAUS M.,
RA KAYTON M., HOUCK K., KISIEL W., SELJELID R., BURNIER J., STERN D.;
RL J. BIOL. CHEM. 269:9774-9782(1994).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE: 93015897.
RA KAO J., RYAN J., BRETT G., CHEN J., SHEN H., FAN Y., GODMAN G.,
RA FAMILLETTI P.C., WANG F., PAN Y.E., STERN D., CLAUS M.;
RL J. BIOL. CHEM. 267:20239-20247(1992).
DR ENML: U10118; G49812; -;
DR MGD: MGI102774; ENAP2.
FT CHAIN 145 310
FT
SQ SEQUENCE 310 AA; 33997 MW; C6DCF8BE CRC32;
POLYPEPTIDE II.
Query Match 49.7%; Score 596; DB 10; Length 310;
Best Local Similarity 54.1%; Pred. No. 4.63e-103;
Matches 86; Conservative 33; Mismatches 35; Indels 5; Gaps 4;
Db 148 IDASRLDIRGICITAKKHPDADSLYEVDVGEAARFTVSGIVNHPLEOMONRAVYL 207
:|||||:|:|:| |||||||||:||||| |||:|:|:|:|:

```

QY 3 VIPSRLDIRVGIITVEKHPDADSLYEKIDVGEAEPRIVSGLVGFVPEKELQDRLV 62
DB 208 LCNLKPAMKMGVLSQAMVNCAS--SPEK-VEILAPNGSVGDRITFDAPP-GEPEKELN 263
QY 63 LCNLKPAMKMGVLSQAMVNCAS--SPEK-VEILAPNGSVGDRITFDAPP-GEPEKELN 122
DB 264 PKKIMEQIDPDHITNAECVATYKADFEVK-GKVCRA 301
QY 123 PKKVEKIQADDFKISECIAOMKQTNFMTKLGISCKS 161

RESULT 2
ID Q20970 PRELIMINARY; PRT; 917 AA.
AC Q20970;
DT 01-NOV-1996 (TREMBLREL, 01, CREATED)
DT 01-NOV-1996 (TREMBLREL, 01, LAST SEQUENCE UPDATE)
DE 01-NOV-1996 (TREMBLREL, 01, LAST ANNOTATION UPDATE)
OS F5883.5.
OC CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; ACOELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RN [1]
RP SEQUENCE FROM N.A.
RA HARRIS B.;
RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRATON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C., MCIMRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C., RIKKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SUSTON J., THIERRY-MING J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOLDMAN P.;
RL NATURE 368:32-38(1994).
EMBL; Z73427; E244446;
SQ SEQUENCE 917 AA; 101713 MW; D148DA58 CRC32;

Query Match 49.7%; Score 595; DB 3; Length 917;
Best Local Similarity 51.5%; Pred. No. 7,72e-103;
Matches 87; Conservative 39; Mismatches 37; Indels 6; Gaps 5;

DB 752 DDTIDVGLDMRGRIITKEKHPDADSLYEKIDVGEAEPRIVSGLVGFVPEKELQDRL 811
QY 1 EEVYIP-SRLDIRVGIITVEKHPDADSLYEKIDVGEAEPRIVSGLVGFVPEKELQDRL 59
DB 812 VVVLNLKPAMKMGVLSQAMVNCAS--SPDK-VEIMEVPADSKPGTPVVCPPYTH-RPDE 867
QY 60 VVVLNLKPAMKMGVLSQAMVNCAS--SPDK-VEIMEVPADSKPGTPVVCPPYTH-RPDE 119
DB 868 QLNPKKKIWTVAEDLKVSAGFAEKMGQPLLGSESKMPTPLRGVH 916
QY 120 ELKPKKRVFEKIQADDFKISECIAOMKQTNFMTKLGISCKSLKGN 167

RESULT 3
ID Q12904 PRELIMINARY; PRT; 312 AA.
AC Q12904;
DT 01-NOV-1996 (TREMBLREL, 01, CREATED)
DT 01-NOV-1996 (TREMBLREL, 01, LAST SEQUENCE UPDATE)
DE 01-NOV-1996 (TREMBLREL, 01, LAST ANNOTATION UPDATE)
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95014290.
RA KAO J., HOCK K., PAN Y., HAHNEL I., LIBUTTI S.K., KAYTON M.L., GRINSCHETT T., CHABOT J., NOWYGRON R., GREENBERG S., KUANG W.J.;

```

```

RA LEUNG D.W., HAYWARD J.R., KISIEL W., HEATH M., BRETT J.,
RA STERN D.M.;
RL J. BIOL. CHEM. 269:25106-25119(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94193665.
RA KAYTON M., HOCK K., KISIEL W., SELJELID R., BURNIER J., STERN D.;
RL J. BIOL. CHEM. 269:9774-9782(1994).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93015897.
RA KAO J., RYAN J., BRETT G., CHEN J., SHEN H., PAN Y., GODMAN G.,
RA FAMILLETI P.C., WANG F., PAN Y.E., STERN D., CLAUS M.;
RL J. BIOL. CHEM. 267:20239-20247(1992).
DR EMBL; U10117; G498910;
FT CHAIN 147 312 ENDOTHELIAL-MONOCYTE ACTIVATING
POLYPEPTIDE II.
SQ SEQUENCE 312 AA; 34326 MW; D012B77F CRC32;

Query Match 49.1%; Score 588; DB 2; Length 312;
Best Local Similarity 53.8%; Pred. No. 2,76e-101;
Matches 84; Conservative 32; Mismatches 35; Indels 5; Gaps 4;

DB 153 SRLDIRVGIITVEKHPDADSLYEKIDVGEAEPRIVSGLVGFVPEKELQDRLV 212
QY 6 SRLDIRVGIITVEKHPDADSLYEKIDVGEAEPRIVSGLVGFVPEKELQDRLV 65
DB 213 LKPKKRGVLSQAMVNCAS--SPEK-TEILAPNGSVGDRITFDAPP-GEPEKELNPK 268
QY 66 LKPKKRGVLSQAMVNCAS--SPEK-TEILAPNGSVGDRITFDAPP-GEPEKELNPK 125
DB 269 KIWEOIOPDLHTNECVATYKADFEVK-GKVCRA 303
QY 126 KVEKIQADDFKISECIAOMKQTNFMTKLGISCKS 161

RESULT 4
ID Q28819 PRELIMINARY; PRT; 658 AA.
AC Q28819;
DT 01-JAN-1998 (TREMBLREL, 05, CREATED)
DT 01-JAN-1998 (TREMBLREL, 05, LAST SEQUENCE UPDATE)
DE 01-JAN-1998 (TREMBLREL, 05, LAST ANNOTATION UPDATE)
OS METHONYL-TRNA SYNTHETASE (METS).
GN AF1453.
OC ARCHAEoglobus FUGIDIS.
OC ARCHAEABACTERIA; EURYARCHAEOTA; ARCHAEoglobales; ARCHAEoglobaceae.
RN [1]
RP SEQUENCE FROM N.A.
RA KLENK H.P., CLAYTON R.A., TOMB J., WHITE O., NELSON K.E., KETCHUM K.A.,
RA DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D., RICHARDSON D.L.,
RA KERLAVAGE A.R., GRAHAM D.E., KRYPIDES N.C., FLEISCHMANN R.D.,
RA DOUGHERTY B.A., LEE N.H., SUTTON G.G., GILL S., KIRKNESS E.F.,
RA REICH C.I., MCNEIL L.K., BADGER J.H., GLODER A., ZHOU L., OVERBECK R.,
RA GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTTERBACK T., COTTON M.D.,
RA SPRIGGS T., ARTIACH P., KAINE B.P., SYKES S.M., SADOW P.W.,
RA D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A., MASON T.M.,
RA OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.;
RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RA KLENK H.P., CLAYTON R.A., TOMB J., WHITE O., NELSON K.E., KETCHUM K.A.,
RA DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D., RICHARDSON D.L.,
RA KERLAVAGE A.R., GRAHAM D.E., KRYPIDES N.C., FLEISCHMANN R.D.,
RA DOUGHERTY B.A., LEE N.H., SUTTON G.G., GILL S., KIRKNESS E.F.,
RA REICH C.I., MCNEIL L.K., BADGER J.H., GLODER A., ZHOU L., OVERBECK R.,
RA GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTTERBACK T., COTTON M.D.,
RA SPRIGGS T., ARTIACH P., KAINE B.P., SYKES S.M., SADOW P.W.,
RA D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A., MASON T.M.,
RA OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.;
RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.

```

DR EMBL: AE001003; G2649118; -
 KW AMINOACYL-TRNA SYNTHASE.
 SQ SEQUENCE 658 AA; 75695 MW; 7761EB99 CRC32;

Query Match 22.0%; Score 263; DB 9; Length 658;
 Best Local Similarity 44.1%; Pred. No. 1,34e-31;

Matches 45; Conservative 19; Mismatches 32; Indels 6; Gaps 5;

DB 561 KIDIRGRVLAKEKVKRSKRLIKLIDIGD-EGQIVSGIAEDYTP-EELEGKLVVLAN 618
 QY 7 RDIIRGKIIIVKHPDADSLYKERIDVGEAEPRIVSGIVQ-FVPEKELODRLVVLCN 65

DB 619 LKPAFMGVESGGMILAEKDG--KAV-LITPEKEVEPGTRV 657
 QY 66 LKPKMRGVESGGMILCAISIEGINROVEPLDPAGSAPGEHV 107

RESULT 5
 ID 033925; PRELIMINARY; PRT; 629 AA.

DT 01-JAN-1998 (TREMBLREL. 05, CREATED)

DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)

DE METHIONYL-TRNA SYNTHETASE.

OS THERMOTOGA MARITIMA.

OC PROKARYOTA; NOT YET CLASSIFIED.

RP STRAIN-MSB8;

RC STRAIN-MSB8;

RA SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.

RL EMBL: U76417; G2645225; -

KM AMINOACYL-TRNA SYNTHASE.

SQ SEQUENCE 629 AA; 73004 MW; 7209A0F4 CRC32;

Query Match 19.7%; Score 236; DB 9; Length 629;
 Best Local Similarity 45.5%; Pred. No. 3,34e-26;

Matches 35; Conservative 21; Mismatches 18; Indels 3; Gaps 3;

DB 531 SKVDRIANVLAEKVPNSRKLRLIIDIG-TEKRIYAGIAEHYR-ELVYGLVYVA 588
 QY 6 SRLDIRVGIITVEKHPDADSLYKERIDVGEAEPRIVSGIV-QFVPEKELODRLVVLC 64

DB 589 NLKPAKMGIESOGML 605
 QY 65 NLKPKMRGVESOGML 81

RESULT 6
 ID 026687; PRELIMINARY; PRT; 651 AA.

AC 026687;

DT 01-JAN-1998 (TREMBLREL. 05, CREATED)

DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)

DE METHIONYL-TRNA SYNTHETASE.

GN MTH587.

OS METHANOBACTERIUM THERMOAUTOTROPHICUM.

OC ARCHAEABACTERIA; EURYARCHAEOTA; METHANOBACTERIALES;

OC METHANOBACTERIACEAE.

RP STRAIN-DELTA H;

RC STRAIN-DELTA H;

RA SMITH D.R.; DOUCETTE-STAMM L.A.; DELOUGHERY C.; LEE H.-M.; DUBOIS J.,

RA ALDEDEDE T.; BASHIRZADER R.; BLAKELY D.; COOK R.; GILBERT K.,

RA HARRISON D.; HOANG L.; KEAGLE P.; LUM W.; POTIER B.; QIU D.,

RA SPADAFORA R.; VICARE R.; WANG Y.; WIERZBOWSKI J.; GIBSON R.,

RA JIWANI N.; CARUSO A.; BUSH D.; SAFER H.; PATWELL D.; PRABHAKAR S.,

RA MCDUGALL S.; SHIMER G.; GOYAL A.; PIETROVSKI S.; CHURCH G.M.,
 RA DANIELS C.J.; MAO J.-I.; RICE P.; NOLLING J.; REEVE J.N.,
 RL J. BACTERIOL. 179:7135-7155(1997).

RP SEQUENCE FROM N.A.

RC STRAIN-DELTA H;

RA SMITH D.R.;

RL SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.

DR EMBL: AE000841; G2621666; -

KM AMINOACYL-TRNA SYNTHASE.

SQ SEQUENCE 651 AA; 74631 MW; C8B64F97 CRC32;

Query Match 18.5%; Score 222; DB 9; Length 651;
 Best Local Similarity 44.7%; Pred. No. 1,88e-23;

Matches 42; Conservative 20; Mismatches 27; Indels 5; Gaps 5;

DB 558 MDIRGVIRSAERIGSDKLLIIDVGEREMO-VVAGIAEKYSP-EDLVERKITVYVL 615
 QY 8 LDIRVGKIIIVKHPDADSLYKERIDVGEAEPRIVSGIVQ-FVPEKELODRLVVLCN 66

DB 616 KPAKLEGVKSEGMLV-ANGESINI-LDPGDAEVG 647
 QY 67 KPQKMRGVESGGMILCAISIEGINROVEPLDPAG 100

RESULT 7
 ID 027056; PRELIMINARY; PRT; 246 AA.

DT 01-JAN-1998 (TREMBLREL. 05, CREATED)

DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)

DE CONSERVED PROTEIN.

GN MTH975.

OS METHANOBACTERIUM THERMOAUTOTROPHICUM.

OC ARCHAEABACTERIA; EURYARCHAEOTA; METHANOBACTERIALES;

OC METHANOBACTERIACEAE.

RP STRAIN-DELTA H;

RC STRAIN-DELTA H;

RA SMITH D.R.; DOUCETTE-STAMM L.A.; DELOUGHERY C.; LEE H.-M.; DUBOIS J.,

RA ALDEDEDE T.; BASHIRZADER R.; BLAKELY D.; COOK R.; GILBERT K.,

RA HARRISON D.; HOANG L.; KEAGLE P.; LUM W.; POTIER B.; QIU D.,

RA SPADAFORA R.; VICARE R.; WANG Y.; WIERZBOWSKI J.; GIBSON R.; JIWANI N.,

RA CARUSO A.; BUSH D.; SAFER H.; PATWELL D.; PRABHAKAR S.; MCDUGALL S.,

RA SHIMER G.; GOYAL A.; PIETROVSKI S.; CHURCH G.M.; DANIELS C.J.,

RA MAO J.-I.; RICE P.; NOLLING J.; REEVE J.N.,

RL J. BACTERIOL. 179:7135-7155(1997).

RP SEQUENCE FROM N.A.

RC STRAIN-DELTA H;

RA SMITH D.R.;

RL SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.

DR EMBL: AE000871; G2622074; -

SQ SEQUENCE 246 AA; 27540 MW; A6538793 CRC32;

Query Match 11.5%; Score 138; DB 9; Length 246;
 Best Local Similarity 36.0%; Pred. No. 5,90e-08;

Matches 31; Conservative 23; Mismatches 23; Indels 9; Gaps 6;

DB 142 VDIINGEVASVGNHVPADLHYCNVAGRSIKVTNDL-D-V-REN--DHVAVAL--LP 194
 QY 8 LDIRVGKIIIVKHPDADSLYKERIDVGEAEPRIVSGIVQ-FVPEKELODRLVVLCN 67

DB 195 PONEGVTSSEGMFL--GVEGVLRDVG 218
 QY 68 POKMRGVESGGMILCAISIEGINROVE 93

RESULT 8
 ID 034943; PRELIMINARY; PRT; 201 AA.

DT 01-JAN-1998 (TREMBLREL. 05, CREATED)

DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)

DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
 DE YTPR.
 GN BACILLUS SUBTILIS.
 OS BACILLUS SUBTILIS.
 CC PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.
 AN [1]
 RP SEQUENCE FROM N.A.
 RA LAPIDUS A., GALLERON N., SOROKIN A., EHRLICH D.;
 RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RA KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALLONI G.,
 AAEVEDO V., BERTERO M.G., BESSIERES P., BOLOTIN A., BORCHERT S.,
 RA BORRIS R., BOUSIER L., BRANS A., BRAUN M., BRIGHELL S.C., BRON S.,
 RA BROUILLET S., BUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,
 RA CHOI S.K., CODANI J.J., CONNERTON I.F., CUMINGS N.J., DANIEL R.A.,
 RA DENIZOT F., DEVINE K.M., DUSTERHOFF A., EHRLICH S.D., EMERSON P.T.,
 RA ENRIAN K.D., ERINGTON J., FABRET C., FERRARI E., FOULGER D.,
 RA FRITZ C., FUJITA M., FUJITA Y., FUMA S., GALLIZI A., GALLERON N.,
 RA GHIM S.Y., GLASER P., GOFEAU A., GOLIGHTLY E.J., GRANDI G.,
 RA GUSEPI G., GUY B.J., HAGA K., HAICH J., HARMOOD C.R., HENAUT A.,
 RA HILBERT H., HOLSAPPEL S., HOSONO S., HULLO M.F., ITAYA M., JONES L.,
 RA JORIS B., KARAMATA D., KASAHARA Y., KLAER-BLANCHARD M., KLEIN C.,
 RA KOBAYASHI I., KOTTER P., KONINGSTEIN G., KROCH S., KUMANO M.,
 RA KORITA K., LAPIDUS A., LARDINOIS S., LAUBER J., LAZAREVIC V.,
 RA LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGUE C.,
 RA MEDINA N., MELLOD R.P., MIZUNO M., MOESTL D., NAKAI S., NOBACK M.,
 RA MOONE D., O'REILLY M., OGAMA K., OGIMARA A., OUDGA B., PARK S.H.,
 RA PARO V., POHL T.M., PORTETELLE D., PORROLIK S., PRESCOTT A.M.,
 RA PRESECAN E., PUOTIC C., PORNELLE B., RAPOPORT G., REY M., REYNOLDS S.,
 RA RIEGER M., RIVOLTA C., ROCHA E., ROCHE B., ROSE M., SADIE Y.,
 RA SAKO T., SCANLAN E., SCHLEICH S., SCHROETER R., SCOFONE F.,
 RA SERIGUCHI J., SEKOWSKA A., SEROR S.J., SEROR P., SHIN B.S., SOLDO B.,
 RA SOROKIN A., TACCONI E., TAKAGI T., TAKAHASHI H., TAKEMARU K.,
 RA TAKERCHI M., TANAKOSHI A., TANAKA T., TERESTRA P., TOGNONI A.,
 RA TOSATO V., UCHITAMA S., VANDENBOL M., VANIER F., VASSAROTTI A.,
 RA VIARI A., WABUT R., WEDLER E., WEDLER H., WEITZENEGGER T.,
 RA WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K., YASUMOTO K., YATA K.,
 RA YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E., YOSHIKAWA H., DANCHIN A.,
 RL NATURE 390:249-256(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RA KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;
 RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: AF008220: G2293213; -
 RA EMBL: Z99119: E1185855; -
 SQ SEQUENCE 201 AA; 21696 MW; E012D585 CRC32;
 Query Match 9.0%; Score 108; DB 9; Length 201;
 Best Local Similarity 35.9%; Pred. No. 4.08e-03;
 Matches 28; Conservative 15; Mismatches 31; Indels 4; Gaps 3;
 Db 98 VGVYSEKHPNADKLVCKVYNGVEETLQIVGAPNDQGVVYAKVAMPGLVIKD 157
 Y 12 VGIITFEKHDPADSLYKIDVGEAEPRVVSGL-VQFVPEKEELDRLLVVLNLT--KP 68
 Db 158 AELRGVSSGMI-CSAKE 174
 Y 69 QKMGVESGMLCASIE 86
 RESULT 9 PRELIMINARY; PRT; 336 AA.
 AC 028306;
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
 DE CONSERVED HYPOTHETICAL PROTEIN.
 GN APL1973.
 OS ARCHAEoglobus fulgidus.
 * ARCHAEABACTERIA; EURYARCHAEOTA; ARCHAEoglobales; ARCHAEoglobaceae.

RN [1]
 RP SEQUENCE FROM N.A.
 RA KLEIN H.P., CLAYTON R.A., TOMB J., WHITE O., NELSON K.E., KETCHUM K.A.,
 RA DODSON R.J., GRINN M., HICKEY E.K., PETERSON J.D., RICHARDSON D.L.,
 RA KERLAVAGE A.R., GRAHAM D.E., KIRPIDS N.C., FLEISCHMANN R.D.,
 RA QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S., KIRNESS E.F.,
 RA DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B., PETERSON S.,
 RA REICH C.I., MCNEIL L.R., BADGER J.H., GLODEK A., ZHOU L., OVERBEER R.,
 RA GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTTERBACK T., COTTON M.D.,
 RA SPRIGGS T., ARTACH P., KAINE B.P., SYKES S.M., SADOW P.W.,
 RA D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A., MASON T.M.,
 RA OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.;
 RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA KLEIN H.P., CLAYTON R.A., TOMB J., WHITE O., NELSON K.E., KETCHUM K.A.,
 RA DODSON R.J., GRINN M., HICKEY E.K., PETERSON J.D., RICHARDSON D.L.,
 RA KERLAVAGE A.R., GRAHAM D.E., KIRPIDS N.C., FLEISCHMANN R.D.,
 RA QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S., KIRNESS E.F.,
 RA DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B., PETERSON S.,
 RA REICH C.I., MCNEIL L.R., BADGER J.H., GLODEK A., ZHOU L., OVERBEER R.,
 RA GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTTERBACK T., COTTON M.D.,
 RA SPRIGGS T., ARTACH P., KAINE B.P., SYKES S.M., SADOW P.W.,
 RA D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A., MASON T.M.,
 RA OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.;
 RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: AE00967: G2648568; -
 SQ SEQUENCE 336 AA; 38113 MW; B5BD46C3 CRC32;
 Query Match 8.3%; Score 100; DB 9; Length 336;
 Best Local Similarity 27.6%; Pred. No. 6.26e-02;
 Matches 16; Conservative 16; Mismatches 26; Indels 0; Gaps 0;
 Db 203 IPIVHAKRIVSIEINPEYKLLKLNKLVNGNIVPLDSQFTPLPGVADRVYM 260
 Y 4 IPSRLIRVGKIVKHPADSLYKIDVGEAEPRVVSGLVQFVPEKEELDRLLV 61
 RESULT 10 PRELIMINARY; PRT; 395 AA.
 AC 052942;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
 DE FLAA, FLAB, FLAD, AND FLAC GENES, COMPLETE COS.
 GN FLAD.
 OS RHIZOBIUM MELILOTI, AND SINORHIZOBIUM MELILOTI.
 CC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;
 OS RHIZOBIACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-RU11/001;
 RA PLATZER J., SCHMITT R.;
 RL SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-RU11/001;
 RX MEDLINE: 97474264.
 RA PLATZER J., STERR W., HAUSMANN M., SCHMITT R.;
 RL J. BACTERIOL. 179:6391-6399(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-RU11/001;
 RA PLATZER J., SCHMITT R.;
 RL SUBMITTED (DEC-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-RU11/001;
 RA SCHMITT R.;
 RL SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: U/6864: G1226187; -
 DR EMBL: I49337: G2275169; -
 SQ SEQUENCE 395 AA; 40955 MW; E9A80973 CRC32;

```

Query Match      8.3%; Score 99; DB 9; Length 395;
Best Local Similarity 30.4%; Pred. No. 8.72e-02;
Matches 21; Conservative 19; Mismatches 24; Indels 5; Gaps 5;

Db 290 INIDITNAQAANDALIRG-DEALED-LISALSALCISIRIMQGEFVSKITDSD 347
OY 29 VERKIDVGEAEPRTVSGLVQFVKELQDLRLVVLNLPKQKMR-GVESQGLM-LCASIE 86
DB 348 SGIGRLVDA 356
OY 87 -GINROVER 94

RESULT 11
ID P74619 PRELIMINARY; PRT; 832 AA.
AC P74619.
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DE 01-FEB-1997 (TREMBLREL. 02, LAST ANNOTATION UPDATE)
DE HYPOTHEICAL 92.9 KD PROTEIN.
OS SYNCHOCYSTIS SP.
OC EUBACTERIA; CYANOBACTERIA; CHROCOCCALES; SYNECHOCYSTIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PCC6803;
RA TABATA S.;
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-PCC6803;
RA TABATA S.;
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-PCC6803;
RA KANeko T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAGATO S., KIMURA T.,
RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARIDO K.,
RA OKUMURA S., SHIMO S., TAKEUCHI C., WADA T., WATANABE A.,
RA YAMADA M., YASUDA M., TABATA S.;
RA DNA RES. 3:109-136(1996).
DR EMBL; D90916; G1653816; -.
KW HYPOTHEICAL PROTEIN.
SQ SEQUENCE 832 AA; 92865 MW; 08554B2A CRC32;

Query Match      8.3%; Score 99; DB 9; Length 832;
Best Local Similarity 32.2%; Pred. No. 8.72e-02;
Matches 19; Conservative 16; Mismatches 21; Indels 3; Gaps 3;

Db 122 KLVTDVNTPAALKS-EOAGNLPRRLNGRS-QVGPLOSLGARGVDVYR-FYQAO 177
OY 58 RLTVVVLNLPKQKMRGVSQGLM-LCASIEGINQVPELDPAGSAGEHVFVKYERQ 116

RESULT 12
ID 022428 PRELIMINARY; PRT; 707 AA.
AC 022428.
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE SIMILAR TO DYAMIN-LIKE PROTEIN.
GN T12EL2.4.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; ACLOMOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE: 94150718.
RA WILSON R., AINSKOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRATON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPER A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., STURSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,

```

```

RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RL NATURE 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA BRADSHAW H., STELLYES L.;
RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA WATERSTON R.;
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U61944; G1397243; -.
DR PROSITE; PS00410; DYAMIN; 1.
SQ SEQUENCE 707 AA; 79335 MW; D151C9FC CRC32;

Query Match      8.2%; Score 98; DB 3; Length 707;
Best Local Similarity 23.1%; Pred. No. 1.21e-01;
Matches 15; Conservative 23; Mismatches 24; Indels 3; Gaps 2;

Db 611 KILTSRQQRVALEIRLIRYFIIVKRNIDSPKAIMLVNFV-RDNIQSLVYRQLY-- 667
OY 7 RLDIRGKTIYVKEKHPDADSLYVERKIDVGEAEPRTVSGLVQFVKELQDLRLVVLNLP 66
DB 668 KPDEM 672
OY 67 KPQKM 71

RESULT 13
ID 041597 PRELIMINARY; PRT; 498 AA.
AC 041597.
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE ADENOSINE TRIPHOSPHATASE (EC 3.6.1.34) (FRAGMENT).
OS TINOSPORA CAFFERA.
OC EUKARYOTA; PLANTAE; EMBRYOIONTA; MAGNOLIOPHYTA; MAGNOLIOPSIDA;
OC MAGNOLIIDEAE; RANUNCULALES; MENISPERMACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RA HOOT S.B., CULHAM A., CRANE P.R.;
RL ANN. MISSOURI BOT. GARDEN 82:194-207(1995).
DR EMBL; L37933; G904161; -.
DR PROSITE; PS00152; ATPASE_ALPHA_BETA; 1.
KW HYDROLASE; HYDROGEN ION TRANSPORT.
FT NON_TER
SQ SEQUENCE 498 AA; 53671 MW; 1F1019D0 CRC32;

Query Match      8.1%; Score 97; DB 8; Length 498;
Best Local Similarity 25.3%; Pred. No. 1.69e-01;
Matches 23; Conservative 27; Mismatches 40; Indels 7; Gaps 5;

Db 204 ERTREGNDLYMKEKSGVINEQNIYESKVALYGGOMNEPPGARMRGVLTALTAEFYRDY 263
OY 19 EKHPPADSLYVKEKIDVGEAEPRTVSGLVQFV--FKELQD-RLTVVVLNLP-KPQKMGV 74
DB 264 NEQDVLL--FDINIFRVOA-GSEVSALIGRXXAVGQ 299
OY 75 ESQGLM-LCASIEGINQVPELDPAGSAGEHVFVKYER 113

RESULT 14
ID P94540 PRELIMINARY; PRT; 804 AA.
AC P94540.
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE PHENYLALANYL-tRNA SYNTHETASE BETA SUBUNIT (EC 6.1.1.20).
GN PHE1.
OS BACILLUS SUBTILIS.
OC PROCARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.
RN [1]

```

RP SEQUENCE FROM N.A.
 RX MEDLINE; 91175935.
 RA BRAKHAGE E.W., WOZNY M., PUTZER H.;
 RL BIOCHIMIE 72:725-734(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA WIPAT A., CARTER N., BRIGHELL C.S., GUY J.B., PIPER K., SANDERS J.,
 RA EMERSON P.T., HARWOOD C.R.;
 RL MICROBIOLOGY 142:3067-3078(1996).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA GOETHL S.F., SCHMID R., WIPAT A., CARTER N.M., EMERSON P.T.,
 RA HARWOOD C.R., NARAHIEL M.A.;
 RL EUR. J. BIOCHEM. 244:59-65(1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RA KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALLONI G.,
 RA ALEVEDO V., BERTERO M.G., BESSIERES P., BOLOTIN A., BORCHERT S.,
 RA BORRIS R., BOUSIER L., BRANS A., BRAUN M., BRINELL S.C., BRON S.,
 RA BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,
 RA CHOI S.K., CODANI J.J., CONNERTON I.F., CUMINGS N.J., DANIEL R.A.,
 RA DENIZOT F., DEVINE K.M., DUSTERHOFT A., EHRLICH S.D., EMERSON P.T.,
 RA ENTIAN K.D., ERRINGTON J., FABRET C., FERRARI E., FOLGER D.,
 RA FIRTZ C., FUJITA M., FUJITA Y., FUMA S., GALIZZI A., GALLERON N.,
 RA GIM S.Y., GLASER P., GOFPEAU A., GOLIGHTLY E.J., GRANDI G.,
 RA GISEPPI G., GUY B.J., HAGA K., HAICH J., HARWOOD C.R., HENAUT A.,
 RA HILBERT H., HOSAPPEL S., HOSONO S., HULLO M.F., ITAYA T., JONES L.,
 RA JORIS B., KARAMATA D., KASAHARA Y., KLAER-BLANCHARD M., KLEIN C.,
 RA KOBAYASHI Y., KOETTER P., KONINGSTEIN G., KROCH S., KUMANO M.,
 RA KORITA K., LAPIDUS A., LARDINOIS S., LAUBER J., LAZAREVIC V.,
 RA LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGUE C.,
 RA MEDINA N., MELLADO R.P., MIZUNO M., MORSTL D., NAKAI S., NOBACK M.,
 RA MOONE D., O'REILLY M., OGAMA K., OGIMARA A., OUDGA B., PARK S.H.,
 RA PARRO V., POHL T.M., PORTETELLE D., PORROLIK S., PRESCOTT A.M.,
 RA PRESECAN E., PUTIC P., PORNELLE B., RAPPOPORT G., REY M., REYNOLDS S.,
 RA RIEGER M., RIVOLTA C., ROCHA E., ROCHE B., ROSE M., SADIE Y.,
 RA SATO T., SCANLAN E., SCHLEICH S., SCHROETER R., SCOPEONE F.,
 RA SERICIGUCHI J., SEKOWSKA A., SEROR S.J., SEROR P., SHIN B.S., SOLDO B.,
 RA SOROKIN A., TACCONI E., TAKAGI T., TAKAHASHI H., TAKEMARU K.,
 RA TOKUCHI M., TANAKOSHI A., TANAKA T., TERPSTRA P., TOGNONI A.,
 RA TOSATO V., UCHIYAMA S., VANDERBOL M., VANNTER P., VASSAROTTI A.,
 RA VARI A., WAMBUUT R., WEDLER E., WEDLER H., WEITENEGGER T.,
 RA WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K., YASUMOTO K., YATA K.,
 RA YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E., YOSHIKAWA H., DANCHIN A.;
 RL NATURE 390:249-256(1997).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RA KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;
 RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -1- CATALYTIC ACTIVITY: ATP + L-PHENYALANINE + TRNA(PHE) = AMP +
 CC DIPHOSPHATE + L-PHENYALANINE-TRNA(PHE).
 DR EMBL; Z75208; E1165325; -;
 DR EMBL; Z99118; E1184112; -;
 KW AMINOCYL-TRNA SYNTHASE; LIGASE.
 SQ SEQUENCE 804 AA; 87945 MW; 25395E1A CRC32;

Query Match 8.1%; Score 97; DB 9; Length 804;
 Best Local Similarity 33.7%; Pred. No. 1.69e-01;
 Matches 29; Conservative 16; Mismatches 33; Indels 8; Gaps 8;

DB 46 VVIGHVLEHQPNAKDLKLVGAEAPVQIGG-APNDKQKAVATGAVLPNGF 104
 QY 10 IRVKGITIVKHPADSLYKIDVGEAEPRIVSGLVQFVKPEE-LQDRIV-VVLC-NL 66
 DB 105 KIKKAKLGEESNGMI-CSLQELGIE 129
 QY 67 K-PQ-KMRGVESQGMGLCASIE-GIN 89

RESULT 15
 ID 029316 PRELIMINARY; PRT; 1036 AA.

AC 029316;
 DT 01-JAN-1998 (TREMBLREL, 05, CREATED)
 DT 01-JAN-1998 (TREMBLREL, 05, LAST SEQUENCE UPDATE)
 DT 01-JAN-1998 (TREMBLREL, 05, LAST ANNOTATION UPDATE)
 DE HYPOTHETICAL 112.4 KD PROTEIN.
 GN AF0946.
 OS ARCHAEOGLOBUS FULGIDUS.
 OC ARCHAEBACTERIA; EURYARCHAEOTA; ARCHAEOGLOBALES; ARCHAEOGLOBACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA KLEIN H.P., CLAYTON R.A., TOMB J., WHITE O., NELSON K.E.,
 RA KETCHUM K.A., DODSON R.J., GWINN M., HICKY E.K., PETERSON J.D.,
 RA RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KIRPIDES N.C.,
 RA FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,
 RA KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,
 RA PETERSON S., REICH C.I., MCNEIL L.R., BADGER J.H., GLOBEK A., ZHOU L.,
 RA OVERBEK R., GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTTERBACK T.,
 RA COTTON M.D., SPRIGGS T., ARTACH P., KATNE B.P., SYKES S.M.,
 RA SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
 RA MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,
 RA VENTER J.C.;
 RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA KLEIN H.P., CLAYTON R.A., TOMB J., WHITE O., NELSON K.E.,
 RA KETCHUM K.A., DODSON R.J., GWINN M., HICKY E.K., PETERSON J.D.,
 RA RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KIRPIDES N.C.,
 RA FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,
 RA KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,
 RA PETERSON S., REICH C.I., MCNEIL L.R., BADGER J.H., GLOBEK A., ZHOU L.,
 RA OVERBEK R., GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTTERBACK T.,
 RA COTTON M.D., SPRIGGS T., ARTACH P., KATNE B.P., SYKES S.M.,
 RA SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
 RA MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,
 RA VENTER J.C.;
 RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; AE001039; G2649670; -;
 KW HYPOTHETICAL PROTEIN.
 SQ SEQUENCE 1036 AA; 112426 MW; F0FDA8F4 CRC32;

Query Match 7.8%; Score 94; DB 9; Length 1036;
 Best Local Similarity 46.7%; Pred. No. 4.45e-01;
 Matches 14; Conservative 9; Mismatches 6; Indels 1; Gaps 1;
 DB 357 PEGYVAKIVGEGELRAEAPVQVYKPK 386
 QY 24 ADSLYEKIDVGEAEPRIVSGLVQFV-PK 52

Search completed: Sat Apr 18 02:04:17 1998
 Job time : 104 secs.

Best Local Similarity 50.0%; Pred. No. 1.59e-75;
Matches 68; Conservative 26; Mismatches 38; Indels 4; Gaps 3;

Db 206 PSALDEVFQKAKIKHPDADSLVSTVDGDEGPRVCGSLKHPFLDMQERIVYV 265
QY 5 PSRLDIRGKIITVEKHPDADSLVEKIDVGEAEPRVSGLVGFVKELQDRLLVVL 63
Db 266 CNLKPVMNRGISTAMVLCGSND--DK-VFEPPEPKSKAGDKYFEGFGDEAPMKOLNP 322
QY 64 CNLKPQKMRGVSOGMLLCASIEGINRQVEPLDPAGAPENHFVNGKQGPDEELKP 123
Db 323 KKKIWEHLQHPFTND 338
QY 124 KKKYFKLQADEKISE 139

RESULT 2
ID YGJH_ECOLI STANDARD; PRT; 110 AA.
AC P42589;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 12.3 KD PROTEIN IN ILEX-EBGR INTERGENIC REGION.
GN YGJH.
OS ESCHERICHIA COLI.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC ENTEROBACTERIACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RA BLATTNER F.R., PLUNKETT G. III, MAYHEW G.F., PERNA N.T., GLASNER F.D.;
RL SUBMITTER (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- SIMILARITY: TO THE C-TERMINAL OF MET TRNA SYNTHETASES AND TO YEAST
CC G4PI.
DR EMBL: U18997; G606012; -
DR EMBL: AE000389; G1789455; -
DR ECOGENE; EG12719; YGJH.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 110 AA; 12315 MW; 65065920 CRC32;

Query Match 25.7%; Score 308; DB 1; Length 110;
Best Local Similarity 50.6%; Pred. No. 1.08e-41;
Matches 41; Conservative 15; Mismatches 24; Indels 1; Gaps 1;

Db 10 ARLEMRVGYKIVEYRHNADKLYVOVDGOKLTQYTS-LVPPYSPEELMGKTIVVLCN 68
QY 6 SRDIDRKGKIITVEKHPDADSLVEKIDVGEAEPRIVSGLVGFVKELQDRLLVVL 65
Db 69 LQKAKMGETSECMILCAETD 89
QY 66 LKPKMRGVSOGMLLCASIE 86

RESULT 3
ID SYM_THERM STANDARD; PRT; 616 AA.
AC P23395;
DT 01-NOV-1991 (REL. 20, CREATED)
DT 01-NOV-1991 (REL. 20, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE METHIONYL-TRNA SYNTHETASE (EC 6.1.1.10) (METHIONINE--TRNA LIGASE)
GN (METS).
OS THERMUS AQUATICUS (SUBSP. THERMOPHILUS).
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;
OC UNCERTAIN.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HB8;
RX MEDLINE: 91131636.
RA NUREKI O., MURAMATSU T., SUZUKI K., KOHDA D., MATSUZAWA H.,
RA OHTA T., MIYAZAWA T., YOKOYAMA S.;
RL J. BIOL. CHEM. 266:3268-3277(1991).
CC -1- CATALYTIC ACTIVITY: ATP + L-METHIONINE + TRNA(MET) = AMP +

CC PYROPHOSPHATE + L-METHIONYL-TRNA(MET).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASES.
DR EMBL: M64273; G155136; -
DR PIR: A39517; SYTMWT.
DR PROSITE: PS00178; AA_TRNA_LIGASE_1; FALSE_NEG.
KW AMINOACYL-TRNA SYNTHETASE; PROTEIN BIOSYNTHESIS; LIGASE; ATP-BINDING.
FT SIMILAR 10 22
FT METAL 297 301 "RMSK" REGION.
FT BINDING 300 300 ATP (BY SIMILARITY).
SQ SEQUENCE 616 AA; 70638 MW; B016F5A CRC32;

Query Match 22.2%; Score 266; DB 1; Length 616;
Best Local Similarity 47.6%; Pred. No. 7.70e-33;
Matches 39; Conservative 22; Mismatches 19; Indels 2; Gaps 2;

Db 520 AKVELRAVEYLAEKHPNADRLVLRISLQNEE-RTVSGIAGWYREELVGKRVLVAN 578
QY 6 SRDIDRKGKIITVEKHPDADSLVEKIDVGEAEPRVSGLVGFVKELQDRLLVVL 65
Db 579 LKPKMRGVSOGML-PADEG 599
QY 66 LKPKMRGVSOGMLLCASIEG 87

RESULT 4
ID SYM_METUA STANDARD; PRT; 651 AA.
AC Q58659;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE METHIONYL-TRNA SYNTHETASE (EC 6.1.1.10) (METHIONINE--TRNA LIGASE)
GN (METS).
GN METS OR MJ1263.
OS METHANOCOCCUS JANNASCHII.
OC ARCHAEABACTERIA; EURYARCHAEOTA; METHANOCOCCALES; METHANOCOCCACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 96337999.
RA BUTT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,
RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,
RA KERLAVAGE A.R., DOUGHERTY B.A., TOMB J.-F., ADAMS M.D., REICH C.I.,
RA OVERBEER R., KIRKGENS E.F., WEINSTOCK K.G., MERRICK J.M., GLADER A.,
RA SCOTT J.L., GEOGHAGEN N.S.M., WEIDMAN J.F., FUHRMANN J.L., NOUYEN D.,
RA UTTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,
RA COTTON M.D., ROBERTS K.M., HURST M.A., KLINE B.P., BORODOVSKY M.,
RA KLENK H.-P., FRASER C.M., SMITH H.O., WOESSE C.R., VENTER J.C.;
RL SCIENCE 273:1058-1073(1996).
CC -1- CATALYTIC ACTIVITY: ATP + L-METHIONINE + TRNA(MET) = AMP +
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASES.
DR EMBL: U67567; G1591898; -
DR PROSITE: PS00178; AA_TRNA_LIGASE_1; FALSE_NEG.
DR TIGR: MJ1263; -
KW AMINOACYL-TRNA SYNTHETASE; PROTEIN BIOSYNTHESIS; LIGASE; ATP-BINDING;
KM METAL-BINDING: ZINC.
FT SIMILAR 8 20
FT METAL 320 324 "HIGH" REGION.
FT METAL 142 142 "RMSK" REGION.
FT METAL 145 145 ZINC (BY SIMILARITY).
FT METAL 154 154 ZINC (BY SIMILARITY).
FT METAL 157 157 ZINC (BY SIMILARITY).
SQ SEQUENCE 651 AA; 75558 MW; 6053C94F CRC32;

Query Match 17.6%; Score 211; DB 1; Length 651;
Best Local Similarity 40.0%; Pred. No. 1.17e-21;
Matches 30; Conservative 20; Mismatches 24; Indels 1; Gaps 1;

Db 553 KILDRGVEVEADIRSKKLKLMVDLD-EKKQIVSGIKGYKPEDLVGKKRIVICNL 611
QY 7 RLDIRGKIITVEKHPDADSLVEKIDVGEAEPRIVSGLVGFVKELQDRLLVVL 66

Db 612 KPAKLGVLSEGMIL 626
 QY 67 KPOKMGVESQGMIL 81

RESULT 5
 ID SYM_HAEIN STANDARD; PRT: 662 AA.
 AC P43828;
 DT 01-NOV-1995 (REL. 32, CREATED)
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
 DE METHIONYL--TRNA SYNTHETASE (EC 6.1.1.10) (METHIONINE--TRNA LIGASE)
 DE (METS).
 GN METG OR H11276.
 OS HAEMOPHILUS INFLUENZAE.
 OC PROKARYOTA: GRACILICUTES; SCOTOBACTERIA: FACULTATIVELY ANAEROBIC RODS;
 OC PASTURELLACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-RD / KW20;
 RX MEDLINE: 95350630.
 RA KERRAVAGE A.R., BULT C.J., TOMB J.-F., DOUGHERTY B.A., KIRKNESS E.F.,
 RA MCKENNEY K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCAYNE J.D.,
 RA SCOTT J.D., SHIRLEY R., LIU L.-I., GIODER A., KELLEY J.M.,
 RA WEIDEMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLUM E., COTTON M.D.,
 RA UETTERACK T.R., HANNA M.C., NGUYEN D.T., SAUDEK D.M., BRANDON R.C.,
 RA FINE L.D., FRITCHMAN J.L., FUHRMANN J.L., GEOGHAGEN N.S.M.,
 RA GNEHM C.L., MCDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,
 RA VENTER J.C.;
 RL SCIENCE 269:496-512(1995).
 CC -1- FUNCTION: IT IS PROBABLY ESSENTIAL FOR CELL SURVIVAL, BEING
 REQUIRED NOT ONLY FOR ELONGATION OF PROTEIN SYNTHESIS BUT ALSO
 FOR THE INITIATION OF ALL MRNA TRANSLATION THROUGH INITIATOR
 CC TRNA(FMET) AMINOACYLATION (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: ATP + L-METHIONINE + TRNA(MET) = AMP +
 PYROPHOSPHATE + L-METHIONYL--TRNA(MET).
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- SIMILARITY: STRONG, TO CYSTEINYL--TRNA SYNTHETASE.
 CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL--TRNA SYNTHETASES.
 DR EMBL: U32807; G1574731; -
 DR PROSITE: PS00178; AA_TRNA_LIGASE_I; 1.
 DR TIGR: H11276; -
 KM AMINOACYL--TRNA SYNTHETASE; PROTEIN BIOSYNTHESIS; LIGASE; ATP-BINDING;
 KM METAL-BINDING; ZINC.
 FT SIMILAR 15 25 "HIGH" REGION.
 FT SIMILAR 331 335 "KMSKS" REGION.
 FT BINDING 334 334 ATP (BY SIMILARITY).
 FT METAL 146 146 ZINC (BY SIMILARITY).
 FT METAL 149 149 ZINC (BY SIMILARITY).
 FT METAL 159 159 ZINC (BY SIMILARITY).
 FT METAL 162 162 ZINC (BY SIMILARITY).
 SQ SEQUENCE 682 AA; 77023 MM; DB6D3990 CRC32;

Query Match 16.7%; Score 200; DB 1; Length 682;
 Best Local Similarity 38.1%; Pred. No. 1.70e-19;
 Matches 32; Conservative 24; Mismatches 24; Indels 4; Gaps 4;

Db 582 AKLDMRVAKVCEKAVPEKSLRPELDLGD-HTRQVSGIKAAVKNKPELEGREVIWVA 640
 QY 6 SRLDIRGKITTVEKHPADSLYEKIDVGEAEPRIVVSGL-VQVPEAELODRLVVLC 64
 Db 641 NLAPRKMKFV-SEGMLISAGTGG 663
 QY 65 NLKPKMR-GVESQGMILCASIEG 87
 RESULT 6
 ID SYM_BACST STANDARD; PRT: 649 AA.
 AC P23920;
 DT 01-MAR-1992 (REL. 21, CREATED)

DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE METHIONYL--TRNA SYNTHETASE (EC 6.1.1.10) (METHIONINE--TRNA LIGASE)
 DE (METS).
 GN METS.
 OS BACILLUS STEAROTHERMOPHILUS.
 OC PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 1518;
 RX MEDLINE: 91305115.
 RA MECHULAM Y., SCHMITT E., PANVET M., SCHMITTER J.-M.,
 RA LAPADAT-TAPOLSKY M., MEINDEL T., DESSEN P., BLANQUET S., FAYAT G.;
 RL NUCLEIC ACIDS RES. 19:3673-3681(1991).
 CC -1- FUNCTION: IT IS PROBABLY ESSENTIAL FOR CELL SURVIVAL, BEING
 REQUIRED NOT ONLY FOR ELONGATION OF PROTEIN SYNTHESIS BUT ALSO
 FOR THE INITIATION OF ALL MRNA TRANSLATION THROUGH INITIATOR
 CC TRNA(FMET) AMINOACYLATION.
 CC -1- CATALYTIC ACTIVITY: ATP + L-METHIONINE + TRNA(MET) = AMP +
 PYROPHOSPHATE + L-METHIONYL--TRNA(MET).
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- SIMILARITY: STRONG, TO CYSTEINYL--TRNA SYNTHETASE.
 CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL--TRNA SYNTHETASES.
 DR EMBL: X57925; G39989; -
 DR PIR: S16682; S16682.
 DR PROSITE: PS00178; AA_TRNA_LIGASE_I; 1.
 KM AMINOACYL--TRNA SYNTHETASE; PROTEIN BIOSYNTHESIS; LIGASE; ATP-BINDING.
 KM SIMILAR 11 23 "HIGH" REGION.
 FT SIMILAR 298 302 "KMSKS" REGION.
 FT BINDING 301 301 ATP (BY SIMILARITY).
 SQ SEQUENCE 649 AA; 74354 MM; 94616A5D CRC32;

Query Match 15.7%; Score 188; DB 1; Length 649;
 Best Local Similarity 37.8%; Pred. No. 3.57e-17;
 Matches 31; Conservative 23; Mismatches 27; Indels 1; Gaps 1;

Db 551 AKVLDRAVAVQPRMKRADLKLQDLG-GEKROYISGAEYKPELIGKKVIVAN 609
 QY 6 SRLDIRGKITTVEKHPADSLYEKIDVGEAEPRIVVSGLVQVPEAELODRLVVLCN 65
 Db 610 LKPAKLGEMSEGMILAGSGG 631
 QY 66 LKPKMGVESQGMILCASIEG 87

RESULT 7
 ID SYM_BACSU STANDARD; PRT: 664 AA.
 AC P37465;
 DT 01-OCT-1994 (REL. 30, CREATED)
 DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
 DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
 DE METHIONYL--TRNA SYNTHETASE (EC 6.1.1.10) (METHIONINE--TRNA LIGASE)
 DE (METS).
 GN METS.
 OS BACILLUS SUBTILIS.
 OC PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RX MEDLINE: 96051385.
 RA OGASAWARA N., NAKAI S., YOSHIKAWA H.;
 RL DNA RES. 1:1-14(1994).
 CC -1- FUNCTION: IT IS PROBABLY ESSENTIAL FOR CELL SURVIVAL, BEING
 REQUIRED NOT ONLY FOR ELONGATION OF PROTEIN SYNTHESIS BUT ALSO
 FOR THE INITIATION OF ALL MRNA TRANSLATION THROUGH INITIATOR
 CC TRNA(FMET) AMINOACYLATION.
 CC -1- CATALYTIC ACTIVITY: ATP + L-METHIONINE + TRNA(MET) = AMP +
 PYROPHOSPHATE + L-METHIONYL--TRNA(MET).
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- SIMILARITY: STRONG, TO CYSTEINYL--TRNA SYNTHETASE.
 CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL--TRNA SYNTHETASES.

DR EMBL: D26185; G467427; -.
 DR SUBLIST: BG10101; METS.
 DR PROSITE: PS00178; AA-TRNA_LIGASE-I; 1.
 KW AMINOACYL-TRNA SYNTHETASE; PROTEIN BIOSYNTHESIS; LIGASE; ATP-BINDING.
 FT SIMILAR 13 25 "HIGH" REGION.
 FT SIMILAR 311 315 "KMSKS" REGION.
 FT BINDING 314 314 ATP (BY SIMILARITY).
 SQ SEQUENCE 664 AA; 76188 MW; E5462587 CRC32;

Query Match 15.18; Score 181; DB 1; Length 664;
 Best Local Similarity 39.58; Pred. No. 7,73e-16;
 Matches 34; Conservative 18; Mismatches 31; Indels 3; Gaps 3;

Db 568 VELRAVEIEAEPEVKRARDLLKLDLGLG-PEKRVVSGIAKHPTPELVKRLVCYT-NL 625
 QY 8 LDIRVGKIIIVKEHPDADSLYVEKIDVGEAEPRVVSGLV-QEVPKELDRLVVLVLCNL 66
 Db 626 KPYKLGELSOGMTILGADGVLYKV 651
 QY 67 KPYKRGVESOGMLLCAISIEGINROV 92

RESULT 8
 ID SYM_HELPY STANDARD; PRT; 650 AA.
 AC P56127;
 DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE METHIONYL-TRNA SYNTHETASE (EC 6.1.1.10) (METHIONINE--TRNA LIGASE)
 DE (METS).
 GN METS OR HP0417.
 OS HELICOBACTER PYLORI (CAMPYLOBACTER PYLORI).
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA;
 OC AEROBIC, MOTILE, HELICAL AND/OR VIBRIOID.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-26695;
 RC MEDLINE: 97394467.
 RX TOMB J.-F., WHITE O., KERLAVAGE A.R., CLAYTON R.A., SUTTON G.G.,
 RA FLEISCHMANN R.D., KETCHUM K.A., KLENK H.-P., GILL S., DOUGHERTY B.A.,
 RA NELSON K., QUACKENBUSH J., ZHOU L., KIRKNESS E.F., PETERSON S.,
 RA LOFTUS B., RICHARDSON D., DODSON R., KHALAK H.G., GLODER A.,
 RA MCKENNEY K., FITZGERALD L.M., LEE N., ADAMS M.D., HICKEY E.K.,
 RA BERG D.E., GOAVALNE J.D., UTERBACK T.R., PETERSON J.D., KELLEY J.M.,
 RA COTTON M.D., WEIDMAN J.M., FUJII C., BOWMAN C., MATTHEY L., WALLIN E.,
 RA HAYES W.S., BORODOVSKY M., KARP P.D., SMITH H.O., FRASER C.M.,
 RA VENTER J.C.;
 RA VENTER J.C.;
 RL NATURE 388:539-547(1997).
 CC -1- FUNCTION: IT IS PROBABLY ESSENTIAL FOR CELL SURVIVAL, BEING
 CC REQUIRED NOT ONLY FOR ELONGATION OF PROTEIN SYNTHESIS BUT ALSO
 CC FOR THE INITIATION OF ALL MRNA TRANSLATION THROUGH INITIATOR
 CC TRNA(FMET) AMINOACYLATION (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: ATP + L-METHIONINE + TRNA(MET) = AMP +
 CC PYROPHOSPHATE + L-METHIONYL-TRNA(MET).
 CC -1- COFACTOR: BINDS A ZINC ION (BY SIMILARITY).
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- SIMILARITY: STRONG, TO CYSTEINYL-TRNA SYNTHETASE.
 CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASES.
 DR PROSITE: PS00178; AA-TRNA_LIGASE-I; FALSE_NEG.
 DR TIGR: HP0417; -.
 KW AMINOACYL-TRNA SYNTHETASE; PROTEIN BIOSYNTHESIS; LIGASE.
 FT SIMILAR 9 21 "HIGH" REGION.
 FT SIMILAR 301 305 "KMSKS" REGION.
 FT BINDING 304 304 ATP (BY SIMILARITY).
 SQ SEQUENCE 650 AA; 74506 MW; CAB14865 CRC32;

Query Match 14.48; Score 173; DB 1; Length 650;
 Best Local Similarity 35.58; Pred. No. 2.49e-14;
 Matches 27; Conservative 23; Mismatches 24; Indels 2; Gaps 2;

Db 552 KYEIKVGLIKAEQRIKSNKILRLKYDVGEGRLQIISGIALDYP-ESLVGQWCVVAN 610
 QY 611 LKPKLGMESGMIT 626
 QY 66 LKPKRGVESOGML 81

QY 7 RDIRVGKIIIVKEHPDADSLYVEKIDVGEAEPRVVSGL-VQFVKELDRLVVLVLCN 65
 Db 611 LKPKLGMESGMIT 626
 QY 66 LKPKRGVESOGML 81

RESULT 9
 ID SYM_ECOLI STANDARD; PRT; 676 AA.
 AC P0059;
 DT 21-JUL-1986 (REL. 01, CREATED)
 DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE METHIONYL-TRNA SYNTHETASE (EC 6.1.1.10) (METHIONINE--TRNA LIGASE)
 DE (METS).
 GN METG.
 OS ESCHERICHIA COLI.
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
 OC ENTEROBACTERIACEAE.
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE: 85054627.
 RA DARDEL F., FAYAT G., BLANQUET S.;
 RL J. BACTERIOL. 160:1115-1122(1984).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE: 91080852.
 RA DARDEL F., PANVER M., FAYAT G.;
 RL MOL. GEN. GENET. 223:121-133(1990).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / BHB2600;
 RA RICHTERICH P., LAKEY N., GRYAN G., JAEHN L., MINTZ L., ROBISON K.,
 RA CHURCH G.M.;
 RL SUBMITTED (OCT-1993) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / M61655;
 RA BLATTNER F.R., PLUNKETT G. III, MAYHEW G.F., PERNA N.T., GLASNER F.D.;
 RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [5]
 RP SEQUENCE OF 1-565 FROM N.A.
 RX MEDLINE: 83079258.
 RA BARKER D.G., EBEL J.-P., JAKES R., BRUTON C.J.;
 RL EUR. J. BIOCHEM. 127:449-457(1982).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
 RX MEDLINE: 82192427.
 RA ZELWER C., RISLER J.L., BRUNIE S.;
 RL J. MOL. BIOL. 155:63-81(1982).
 RN [7]
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
 RX MEDLINE: 91073404.
 RA BRUNIE S., ZELWER C., RISLER J.L.;
 RL J. MOL. BIOL. 216:411-424(1990).
 RN [8]
 RP STRUCTURE BY NMR OF 138-163.
 RX MEDLINE: 93294859.
 RA FOURRY D., DARDEL F., BLANQUET S.;
 RL J. MOL. BIOL. 231:1078-1089(1993).
 RN [9]
 RP ACTIVE SITE MAPPING.
 RX MEDLINE: 91084494.
 RA HOUTONDUIT C., SCHMITTER J.-M., BEAUVALLLET C., BLANQUET S.;
 RL BIOCHEMISTRY 29:8190-8198(1990).
 RN [10]
 RP MUTAGENESIS.
 RX MEDLINE: 93070503.
 RA FOURRY D., MECHULAM Y., BRUNIE S., BLANQUET S., FAYAT G.;
 RL FEBS LETT. 292:259-263(1991).
 RN [11]
 RP MUTAGENESIS OF ZINC-LINGANDS.

ID STFB_HELPY STANDARD; PRT: 764 AA.
 AC P56145;
 DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE PHENYLAANYL-TRNA SYNTHETASE BETA CHAIN (EC 6.1.1.20) (PHENYLAANYNE--
 TRNA LIGASE BETA CHAIN) (PHERS).
 GN PHE1 OR HP0402.
 OS HELICOBACTER PYLORI (CAMPILOBACTER PYLORI).
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA;
 CC AEROBIC, MOTILE, HELICAL AND/OR VIBRIOID.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=26695;
 RX MEDLINE; 97394467.
 RA TOMB J.-F., WHITE O., KERLAVAGE A.R., CLAYTON R.A., SUTTON G.G.,
 RA FLEISCHMANN R.D., KETCHUM K.A., KLENK H.-P., GILL S., DOUGHERTY B.A.,
 RA NELSON K., QUACKENBUSH J., ZHOU L., KIRKNESS E.F., PETERSON S.,
 RA LOFTUS B., RICHARDSON D., DODSON R., KHALAK H.G., GLODER A.,
 RA MCKENNEY K., FITZGERALD L.M., LEE N., ADAMS M.D., HICKEY E.K.,
 RA BERG D.E., GOCAYNE J.D., UTTERBACK T.R., PETERSON J.D., KELLEY J.M.,
 RA COTTON M.D., WEIDMAN J.M., FUJII C., BOWMAN C., WATTEY L., WALLIN E.,
 RA HAYES W.S., BORODOVSKY M., KARP P.D., SMITH H.O., FRASER C.M.,
 RA VENTER J.C.; 339-547(1997).
 RL NATURE 388:339-547(1997).
 CC -1- CATALYTIC ACTIVITY: ATP + L-PHENYLAANYNE + TRNA(PHE) = AMP +
 CC PYROPHOSPHATE + L-PHENYLAANYL-TRNA(PHE).
 CC -1- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA CHAINS
 CC (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 DR TIGR: HP0402;
 KW AMINOACYL-TRNA SYNTHETASE; PROTEIN BIOSYNTHESIS; LIGASE; ATP-BINDING.
 SQ SEQUENCE 764 AA; 85135 MW; 93C8EB3B CRC32;
 Query Match 10.7%; Score 128; DB 1; Length 764;
 Best Local Similarity 36.8%; Pred. No. 2.51e-06;
 Matches 32; Conservative 18; Mismatches 29; Indels 8; Gaps 6;
 Db 39 IAPK-NNVVGKILKAPKNAEKLVCVDGKEVLQIVCGAKNAPQFVP-VALNGAL 96
 Qy 4 IPSRLDLRVGKIIVKHPDADSLYVERKIDGAEPRIVSGL-V---QVFKKELQDRL 59
 Db 97 IGSTIAKTE-LRGVESHGMI-CSSIE 121
 Qy 60 VVVLCLNKPQKMGVSGMGLCASIE 86
 RESULT 13
 ID STFB_ECOLI STANDARD; PRT: 795 AA.
 AC P07395; Q59407;
 DT 01-APR-1988 (REL. 07, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE PHENYLAANYL-TRNA SYNTHETASE BETA CHAIN (EC 6.1.1.20) (PHENYLAANYNE--
 TRNA LIGASE BETA CHAIN) (PHERS).
 GN PHE1.
 OS ESCHERICHIA COLI.
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
 CC ENTEROBACTERIACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 85261100.
 RA MECHOLMAN Y., FAYAT G., BLANQUET S.;
 RL J. BACTERIOL. 163:787-791(1985).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MILLER H.I.;
 RL SUBMITTED (NOV-1986) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RA BLATTNER F.R., PLUCKETT G. III, MAYHEW G.F., PERNA N.T., GLASNER F.D.;
 RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.

RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RA AIBA H., BABA T., FUJITA K., HAYASHI K., HONJO A., HORIUCHI T.,
 RA IKEMOTO K., INADA T., ISONO K., ISONO S., ITOH T., KAKI K., KASAI H.,
 RA KASHIMOTO K., KIM S., KIMURA S., KITAGAWA M., KITAKAWA M., MAKINO K.,
 RA MASUDA S., MIKI T., MIZOBUCHI K., MORI H., MOTOMURA K., NAKAMURA Y.,
 RA NASHIMOTO H., NISHIO Y., OSHIMA T., SAITO N., SAMPET G., SEKI Y.,
 RA TAGAMI H., TAKEMOTO K., WADA C., YAMAMOTO Y., YANO M.;
 RA SUBMITTED (DEC-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -1- CATALYTIC ACTIVITY: ATP + L-PHENYLAANYNE + TRNA(PHE) = AMP +
 CC PYROPHOSPHATE + L-PHENYLAANYL-TRNA(PHE).
 CC -1- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA CHAINS.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 DR EMBL: V00291; G43071; -;
 DR EMBL: K02844; G146346; -;
 DR EMBL: A0000266; G1788006; -;
 DR EMBL: D90814; G1742806; -;
 DR EMBL: D90813; G1742793; -;
 DR PIR: B23099; SECEFB.
 DR SWISS-2DPAGE: P07395; COLI.
 DR ECO2DPAGE: D094.0; 6TH EDITION.
 DR ECOGENE; EG10710; PHE1.
 KW AMINOACYL-TRNA SYNTHETASE; PROTEIN BIOSYNTHESIS; LIGASE; ATP-BINDING.
 FT CONFLICT 93 93 A -> T (IN REF. 1).
 FT CONFLICT 141 142 AP -> VR (IN REF. 1).
 FT CONFLICT 186 189 PLVO -> AAGN (IN REF. 1).
 FT CONFLICT 481 481 A -> R (IN REF. 1).
 FT CONFLICT 698 698 R -> G (IN REF. 1).
 SQ SEQUENCE 795 AA; 87378 MW; 5E5C3A3C CRC32;
 Query Match 8.8%; Score 106; DB 1; Length 795;
 Best Local Similarity 29.2%; Pred. No. 7.86e-03;
 Matches 33; Conservative 22; Mismatches 51; Indels 7; Gaps 6;
 Db 36 EPVAGSHGVYGEVEVECAQHPNADKLRTKVNVGGRLLDIYCGADNCGGLRVAVAT 95
 Qy 1 EEVIPSRLDLRVGKIIVKHPDADSLYVERKIDGAEPRIVSGLQVFKKELQDRL 59
 Qy 60 VVVLCLN-LKPQKMGVSGMGLCASIE-GI-NRQVEPLDPAGSAPGEHV 107

RESULT 14
 ID STFB_MYCPN STANDARD; PRT: 805 AA.
 AC P75563;
 DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE PHENYLAANYL-TRNA SYNTHETASE BETA CHAIN (EC 6.1.1.20) (PHENYLAANYNE--
 TRNA LIGASE BETA CHAIN) (PHERS).
 GN PHE1.
 OS MYCOPLASMA PNEUMONIAE.
 OC PROKARYOTA; TENERICUTES; MOLICUTES; MYCOPLASMA; MYCOPLASMATALES;
 CC MYCOPLASMATACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 29342 / M129;
 RX MEDLINE; 97105885.
 RA HIMMELREICH R., HILBERT H., PLACENS H., PIRKL E., LI B.-C.,
 RA HERRMANN R.;
 RL NUCLEIC ACIDS RES. 24:4420-4449(1996).
 CC -1- CATALYTIC ACTIVITY: ATP + L-PHENYLAANYNE + TRNA(PHE) = AMP +
 CC PYROPHOSPHATE + L-PHENYLAANYL-TRNA(PHE).
 CC -1- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA CHAINS
 CC (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 DR EMBL: A000006; G1673698; -;
 KW AMINOACYL-TRNA SYNTHETASE; PROTEIN BIOSYNTHESIS; LIGASE; ATP-BINDING.
 SQ SEQUENCE 805 AA; 91713 MW; F4F657DC CRC32;
 Query Match 8.8%; Score 105; DB 1; Length 805;

!!SEQUENCE_LIST 1.0

(Nucleotide) FASTA of: seq1a from: 1 to: 504 April 28, 1998 08:40

TO: geneseqn: * Sequences: 159,646 Symbols: 57,748,920 Word Size: 6

Sequences too short to analyze: 10 (42 symbols)

Databases searched: Geneseq-NA, Release 30.0, Released on 26Jan1998, Formatted on 10Mar1998

Searching with both strands of the query.

Scoring matrix: GenRunData:fastadna.cmp

Constant pamfactor used

Gap creation penalty: 16 Gap extension penalty: 4

Histogram Key:

Each histogram symbol represents 787 search set sequences
Each inset symbol represents 3 search set sequences
z-scores computed from opt scoresz-score Obs exp
(=) (*)

```

< 20 518 0 : *
22 26 0 : *
24 9 0 : *
26 36 6 : *
28 76 64 : *
30 197 388 : *
32 596 1500 : *
34 1692 4069 : *
36 4018 8356 : *
38 7410 13809 : *
40 11127 19263 : *
42 19191 23546 : *
44 32719 25974 : *
46 39908 26455 : *
48 47177 25328 : *
50 25910 23112 : *
52 27384 20319 : *
54 13730 17356 : *
56 10401 14498 : *
58 9391 11902 : *
60 6335 9641 : *
62 5619 7730 : *
64 4236 6147 : *
66 3167 4859 : *
68 2586 3822 : *
70 1919 2995 : *
72 1629 2340 : *
74 1187 1825 : *
76 899 1420 : *
78 615 1104 : *
80 590 857 : *
82 470 656 : *
84 312 519 : *
86 228 402 : *
88 170 311 : *
90 169 241 : *
92 90 186 : *
94 120 144 : *
96 58 111 : *
98 47 86 : *
100 36 67 : *
102 33 52 : *
104 26 40 : *
106 30 31 : *
108 28 24 : *
110 12 19 : *

```

```

112 9 14 : *
114 0 11 : *
116 0 9 : *
118 11 7 : *
>120 20 5 : *

```

Results sorted and z-values calculated from opt score
1968 scores saved that exceeded 81
117963 optimizations performed

Joining threshold: 51, optimization threshold: 36, opt. width: 16

The best scores are:

```

GENESEQN:T49312      Begin: 94 End: 597      Initl Initn opt z-sc E(281635)..
! Human endothelial-monocyte activati... 2520 2520 2520 2556.0 0
GENESEQN:Q86718      Begin: 506 End: 746
! Mouse EMAP111 CDNA... New endothelia... 424 524 449 449.6 3.3e-18
GENESEQN:Q94113      Begin: 242 End: 389      Strand: -
! PML2 genomic DNA... Thrombopoietin ... 72 72 136 132.1 1.7
\\End of List

```

seq1a
GENESEQN:T49312

```

ID T49312 standard; cDNA; 636 BP.
AC T49312;
DF 04-JUL-1997 (first entry)
DE Human endothelial-monocyte activating polypeptide III encoding cDNA.
KW Endothelial-monocyte activating polypeptide; EMAP 111; human;
   neoplasia; ss...

```

SCORES Initl: 2520 Initn: 2520 Opt: 2520 z-score: 2556.0 E(): 0
100.0% Identity in 504 bp overlap

```

seq1a      GAGAGATCATCCATCCCGCGCTGGATATC
T49312      GGCCTAAGGCGCTGCCAAGATTTCAGAACACAGAGGATCCATCCCGCGCTGGATATC
              70      80      90      100      110      120
seq1a      CGTGTGGGAAATCATCATCTGTGAGAGACACCCAGATGCGAGACGCTTATGTATAG
T49312      CGTGTGGGAAATATATCATCTGTGAGAGACACCCAGATGCGAGACGCTTATGTATAG
              130      140      150      160      170      180
seq1a      AAGATTGACGTGGGGAGAGCTGAACACGAGCTGTGTGAGCGGCGCTGTACAGTTGCTG
T49312      AAGATTGACGTGGGGAGAGCTGAACACGAGCTGTGTGAGCGGCGCTGTACAGTTGCTG
              190      200      210      220      230      240
seq1a      CCAGAGGAGAACTGACAGACGAGCTGTGTGAGCTGTGTGAGCTGTGTGAGCTGTGTGAG
T49312      CCAGAGGAGAACTGACAGACGAGCTGTGTGAGCTGTGTGAGCTGTGTGAGCTGTGTGAG
              250      260      270      280      290      300
seq1a      ATGAGAGAGTGCAGTCCCAAGGCGATGCTGTGTGAGCTGTGTGAGCTGTGTGAGCTGTGTGAG
T49312      ATGAGAGAGTGCAGTCCCAAGGCGATGCTGTGTGAGCTGTGTGAGCTGTGTGAGCTGTGTGAG
              310      320      330      340      350      360
seq1a      CAGGTGAACCTGTGAGACCTTCGCGAGGCTCTGCTGCTGTGAGCTGTGTGAGCTGTGTGAG
T49312      CAGGTGAACCTGTGAGACCTTCGCGAGGCTCTGCTGCTGTGAGCTGTGTGAGCTGTGTGAG
              370      380      390      400      410      420
seq1a      GCGTATGAAGAGGCGCAACCAAGATGAGAGCTCAAGGCCAAGAGAAAGCTTGTGAGAG

```

[illegible]

SCORES Initl: 424 Initn: 524 Opt: 449 z-score: 449.6 E(): 3.3e-18
65.1% Identity in 241 bp overlap

[illegible][illegible]

```

i CPU time used:
i Database scan: 0:10:07.9
i Post-scan processing: 0:00:00.3
i Total CPU time: 0:10:08.2
i Output file: seqa.fasta
i _RNA_SEQUENCE 1.0
ID T49312 standard; cDNA, 636 BP.
AC T49312;
DT 04-JUN-1997 (first entry)
DE Human endothelial-monocyte activating polypeptide III encoding cDNA.
KW Endothelial-monocyte activating polypeptide; EMAP III; human;
KW neoplasia; ss.
OS Synthetic.
FH Key
FT CDS Location/Qualifiers
FT CDS 94..600
FT /product= EMAP_III
PN M09640719-A1.
PD 19-DEC-1996
PF 07-JUN-1995; U07328.
PR 07-JUN-1995; WO-U07328.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Coleman TA, Olsen HS, Rosen CA;
PI WPI: 97-052219/05.
DR P-PSDB; M06596.
FT New isolated endothelial-monocyte activating polypeptide III - used
PT to develop prods. for the diagnosis and treatment of conditions
PT involving EMAP III, e.g. Neoplasia
PS Claim 5: Page 37: 50pp; English
CC The present sequence encodes the 168 amino acids of the novel
CC polypeptide endothelial-monocyte activating polypeptide II (EMAP III)
CC The product can be used for the diagnosis and treatment of conditions
CC involving abnormal expression of EMAP III. In particular, EMAP III can
CC be used to regress neoplasia such as in tumours and cancers.
CC EMAP III shows a high degree of homology to EMAP II with 60%
CC identity and 75% similarity over a 150 amino acid stretch.
SQ Sequence 636 BP; 172 A; 166 C; 178 G; 120 T;

```

```

T49312 Length: 636 April 28, 1998 12:50 Type: N Check: 4493 ..
1 TACCCCTGCC CTGAAAAAC TGGCCAGCCG TGCCTACCCA GATCCCTCAA
51 AGCAGAACCC AATGGCCAAA GGCCTGCCAA GAATTCAGAA CCAGAGGAGG
101 TCATCCCATC CCGGCTGGAT ATCCGTGTGG GGAATATCAT CACTGTGGAG
151 AAGACCCAG ATGCAGACAG CCTGTATGTA GAGAAGATTG ACGTGGGGGA
201 AGCTGAACCA CGGACTGTGG TGAGCGGCTT GGTACAGTTC GTGCCCAAGG
251 AGGAAGTCCA GAGACGGCTG GTAGTGTGTC TGTGCAACCT GAAACCCGAG
301 AAGATGAGAG GAGTGAAGTC CCAAGGCATG CTTGTGTGTG CTTGATAGTA
351 AGGATAAAC CGCCAGTTGG AACCTGTGGA CCTCCGGGCA GGCCTGTGTC
401 CTGTGAGCA CGTGTGTGTG AAGGCTATG AAAAGGGCCA ACCAGATGAG
451 GAGCTCAAGC CCAAGAGAA AGCTTCGAG AAGTTCGAGG CTGACTTCAA
501 AATTTCTGAG GAGTGCATCG CACAGTGAA GCAAAACCAAC TTCATGACCA
551 AGCTGGCTC CATTTCTGT ATATGCTGA AAGGGGGGAA CATTAGCTAG
601 CCAGCCACAG ATCTCCCCC CTCTCTCCAC CACTGA

!!NA_SEQUENCE 1.0
ID 086718 standard; cDNA; 1086 BP.
AC 086718;
DT 29-SEP-1995 (first entry)
DE Mouse EMAPII cDNA.
KW EMAPII; endothelial monocyte activating polypeptide II; chemotaxis;
KW inflammation; tissue factor; tumor; cancer; therapy; metha;
OS Mus musculus.
FH Key Location/Qualifiers
FT CDS 64..996
FT /*tag- a
PN W09509180-A.
PD 06-APR-1995.
PE 29-SEP-1994; U11085.
PR 29-SEP-1993; US-129456.
PA (UYCO ) UNIV COLUMBIA NEW YORK.
PI Clauss M, Rao J, Kayton M, Libutti SK, Stern DM;
DR WPI: 95-147389/19.
DR P-PSDB; R72577.
PT New endothelial monocyte activating polypeptide II - induces
PT chemotaxis, inflammation and tissue factor, useful for treating
PT tumours, also related antibodies, DNA and active fragments
PS Disclosure: Fig.4; 180pp; English.
CC A mouse metha sarcoma cDNA library was screened with a probe
CC based on the N-terminal sequence of mouse EMAPII. Overlapping
CC clones were combined to obtain a contiguous full-length sequence
CC (given in 086718) encoding a 33 kDa protein (R72577).
CC Recombinant EMAPII was expressed in E. coli.
CC Sequence 1086 BP; 346 A; 212 C; 297 G; 231 T;
SO

086718 Length: 1086 April 28, 1998 12:50 Type: N Check: 3316 ..
1 GAGGCTGTC AAGAGCTGCG GTTGGGTCAC CGCTTCATCT TTTCTGCGC
51 ATTCTGGGGA AAGATGCCAA CGAATGATGC TTTCTGANG AGGCTGAGC
101 AGAAGGTGTC AGAGCGCGAT CAGATCATCG AATATCTCAA GCAGCAGGTT
151 GCTCTTCTTA AGGAGAAAGC AATTTTCAG GCAACAATA GAGAAGAAA
201 GAACTTGA GTTGAATG CTAACTGAA AAAAGAAATA GAAGAGCTAA
251 AGCAAGACT GATTCTGCA GAATTCATA ACGAGTGA GCAAGTGGT

```

```

301 GTTCGATTGA GTACTCCACT GCAGACGAAC TGTACTGCTT CTGAAGTGT
351 GGTGAGTCT CCATCACTAG CAACCAACCC CTCTCTCTCT ACAAAAAGC
401 AGATCAAGC GGGAGAGAA AAGAGGTGA AAGAGAGAC TGAAGAGAA
451 GGAGAGAAA AGGAGAGCA GCAGTGGCA GCAGCAAGTA CTGACTCCA
501 GCTATCCAC GCATCGCTC TGAATCTTG AATTGGTGT ATTGTACTG
551 CCAGAGACA CCTGATGA GATTCAGTGT ATGTGAGAGA AGTAGATGTG
601 GGAGAGAGAG CCCGCGCAC GGTGTCAGC GGGGTGTGTA ATCATGTTCC
651 TCTAGAACG ATGCAATC GTATGTGT TTTACTCTCT AATGTAGAC
701 CTGCAAGAT GCGGGGAGT CTGTCTCAAG CCATGTGAT GTGTGCCAGT
751 TCACCAGAG AAGTGAGAT TGTGGCCCT CCACAGGGT CCGTCTGTG
801 GGACGAATT ACTTTGATG CTTTCTCTG AGAGCTTAC AAGAGCTAA
851 ACCCTAGAA GAAGATCTG GAGCAGATCC AGCCTGACT GCACACCAAT
901 GGTAGGTGT TGCCACATA CAAGGAGCT CCTTTGAG TGAAGGGAA
951 GGGATTGTC AGAGCCAAA CCATGGCCA TAGTGAATT AATATGTC
1001 TCTGTACTG AAGAGCATG GCGAAACTT AATAACAATA AAGAGAAGT
1051 TGTATATCAC TTACATATA AAAAAAAAAA AAAAAA

!!NA_SEQUENCE 1.0
ID 094113 standard; DNA; 1014 BP.
AC 094113;
DT 27-FEB-1996 (first entry)
DE PML2 genomic DNA.
KW Human; thrombopoietin; TPO; mpl ligand; hml; fragment polypeptide;
KW megakaryocytopoietic cytokine receptor; thrombopoietic signal;
KW EPO-domain fragment; erythropoietin; hepo; haematopoietic cell;
KW megakaryocyte; thrombocytopenia; myeloproliferative disease;
KW inflammatory thrombocytosis; iron deficiency; Epo; platelet;
KW red blood cell; progenitor; hml-2; ss.
OS Sus scrofa.
FH Key Location/Qualifiers
FT CDS 1..987
FT /*tag- a
FT /product- PML2
PN GB2285446-A.
PD 12-JUL-1995.
PE 21-DEC-1994; 025831.
PR 03-JAN-1994; US-176553.
PR 21-JAN-1994; US-185607.
PR 15-FEB-1994; US-196689.
PR 04-APR-1994; US-223263.
PR 25-MAY-1994; US-249376.
PR 02-DEC-1994; US-348658.
PR 02-DEC-1994; US-348657.
PA (GETH ) GENENTECH INC.
PI Eaton DL, de Sauvage FJ;
DR WPI: 95-234018/31.
DR P-PSDB; R76170.
PT Thrombopoietin polypeptide, ligand for mpl cytokine receptor -
PT useful for treating thrombocytopenia and related diseases
PS Example 13; Fig 21; 192pp; English.
CC This sequence represents a genomic clone encoding an isoform of porcine
CC thrombopoietin (TPO), also known as mpl ligand (PML), pml2. This
CC sequence was isolated by RACE PCR using the primer sequences given in
CC 094140-43 and 094128. The genomic sequence was isolated from a porcine
CC genomic library in EMBL3 by screening with pr45. Two forms of pml were
CC isolated in this manner. The first was a full length clone (see also
CC 094112) and the second encoded a protein with a four amino acid deletion

```

CC (pML2). Comparison of the sequences of pML and pML2 shows that the
CC latter form is identical except for the deletion of the tetrapeptide QLRP
CC corresponding to residues 111-114. The four amino acid deletions
CC observed in murine, human and porcine ML cDNA occur at precisely the
CC same position in the predicted proteins (see also Q94110). pML binds to
CC mPl which is a megakaryocytopoietic cytokine receptor which may transduce
CC a thrombopoietic signal.
SQ Sequence 1014 BP; 216 A; 365 C; 220 G; 213 T;

094113 Length: 1014 April 28, 1998 12:50 Type: N Check: 5829 ..

```
1 AGCCCGGCTC CTCCTGCTG TGACCCCGCA CTCCTAATA AACTGCTTCG
51 TGACTCCCAT GTCCTTCAG GCAGACTGAG CCAAGTGCCA GACATTAAC
101 CTTGTCCAC ACCTGTCTG CTGCCTGCTG TGGACTTAC CTGGAGAA
151 TGGAAACC AGACGGAGCA GACAAAGCA CAGATGTCC TGGAGCCAC
201 AACCTTCTG CTGAGGAGAG TGATGACAGC ACGGGAGCAA GTGGAGCCCC
251 CTGGCCTCTC ATCCCTGCTG GTGCAGCTT CTGGACAGGT TCGCCTCTC
301 CTGGGGGCC TGCAGGACT CCTTGAATG CAGGGAAGA CCACAGCTCA
351 CAAGGATCCC AGTGCCATCT TCCTGAAGT CCAACAAGTG CTCGAGGAA
401 AGGTGGCTT CCGCTCTCT GTAGTGGGG CCTCCCTCTG TGCCAAGAG
451 GCCCCACCG CCATAGCTGT CCGGAGCAGC ACCTCTCCAT TCCACACACT
501 GAACAAGCTC CCAACAGGA CCTGTGATT GTTGAGACA AACTCCAGTA
551 TCTAGCCAG AACTACTGC TCTGATTTT TCAAGAGCT GCAGGCATTC
601 AGAGCCAGA TTCCTGGTCT GCTGACCAG ACCTCCAGT CCCTAGACCA
651 AATCCCTGGA CACAGNATG GGACACAGG ACCCTGAGT GGAATTCATG
701 GACTTTTCC TGGACCCCA CCGGGGGCC TCGAGCTCC AGACATTCCT
751 CCAGCAACTT CAGGATGGG CTCCGGGCA ACTTACCTCC AGCCTGAGA
801 GTCTCTTCC CCAGTCACC CTTCTCTGG ACGATACACT CTCTCTCTC
851 CTCACCCAC CTCGCCCTCC CCCAGATCC AGCTCCAGCC TCTGCTTCT
901 GACCCCTCTG CGATCAGAC CAACTCTACC AGTCTCTTC TATTGACAG
951 TCACCTCAT TTCAGAACG TGTTCAGAG AGAGTAAGT GCTCAGACCC
1001 TGCCAACCTC AGCA
```



```

GENESQ:R10727      Begin: 55      End: 135
I (Glu108Asp109Glu110)IL-3.. New int...
GENESQ:R14839      Begin: 115      End: 192
I Protein deduced from human Glioblas...
GENESQ:R78367      Begin: 43      End: 122
I Human IL-3 mutant for multi-lineage...
GENESQ:R56082      Begin: 44      End: 123
I Human interleukin-3 mutleln (pMon13)...
GENESQ:R72583      Begin: 2      End: 15
I EMAP11 active peptide.. New endoth...
GENESQ:R56093      Begin: 31      End: 110
I Human interleukin-3 mutleln (pMon13)...
GENESQ:R56068      Begin: 31      End: 110
I Human interleukin-3 mutleln (pMon13)...
GENESQ:R56092      Begin: 31      End: 110
I Human interleukin-3 mutleln (pMon13)...
GENESQ:R56101      Begin: 31      End: 110
I Human interleukin-3 mutleln (pMon13)...
GENESQ:R78353      Begin: 31      End: 110
I Human IL-3 mutant for multi-lineage...
GENESQ:R56100      Begin: 31      End: 110
I Human interleukin-3 mutleln (pMon13)...
GENESQ:R56104      Begin: 31      End: 110
I Human interleukin-3 mutleln (pMon13)...
GENESQ:R56094      Begin: 31      End: 110
I Human interleukin-3 mutleln (pMon13)...
GENESQ:R22820      Begin: 44      End: 124
I K64, P83, M88, K113, E132, M127, E1...
GENESQ:R56139      Begin: 31      End: 110
I Human interleukin-3 mutleln (pMon13)...
GENESQ:W13514      Begin: 6      End: 97
I Interleukin-3 receptor agonist pMON...
GENESQ:R56141      Begin: 31      End: 88
I Human interleukin-3 mutleln (pMon13)...
GENESQ:R36739      Begin: 15      End: 60
I Human bone morphogenetic protein BM...
\\End of List

```

```

pep2
GENESQ:W06596
ID W06596 standard; Protein: 168 AA.
AC W06596;
DT 04-JUL-1997 (first entry)
DE Human endothelial-monocyte activating polypeptide III.
KW Endothelial-monocyte activating polypeptide; EMAP III; human;
KW neoplasta...

```

```

SCORES      Init1: 1111      Initn: 1111      Opt: 1111 z-score: 1640.3 E(): 0
Smith-Waterman score: 1111; 100.0% identity in 168 aa overlap

```

```

pep2      10      20      30      40      50      60
EEVPSRLDIRVGKIIIVKEHPDADSLYVEKIDVGEAEPTTVVSGLVQFVKPELDRLY
W06596      10      20      30      40      50      60
EEVPSRLDIRVGKIIIVKEHPDADSLYVEKIDVGEAEPTTVVSGLVQFVKPELDRLY
VVLNLKPKQMRGVSOGMLCASISGIRNQEPLDPDPASAPGEHVFVYKGGQDDE
pep2      70      80      90      100      110      120
VVLNLKPKQMRGVSOGMLCASISGIRNQEPLDPDPASAPGEHVFVYKGGQDDE
W06596      70      80      90      100      110      120
VVLNLKPKQMRGVSOGMLCASISGIRNQEPLDPDPASAPGEHVFVYKGGQDDE

```

```

pep2      130      140      150      160
LKPKKVFEKLQADFKIISSECIAGWKQTNMTKLSISCSKLGKNGIS
W06596      130      140      150      160
LKPKKVFEKLQADFKIISSECIAGWKQTNMTKLSISCSKLGKNGIS
GENESQ:W14561

```

```

ID W14561 standard; protein: 166 AA.
AC W14561;
DT 01-DEC-1997 (first entry)
DE Endothelial monocyte activating polypeptide II.
KW EMAP-II; inhibitor; endothelial cell formation; blood vessel;
KW retinopathy; tumour; subcutaneous; intraperitoneal; intravenous;

```

```

SCORES      Init1: 401      Initn: 546      Opt: 585 z-score: 865.4 E(): 0
Smith-Waterman score: 585; 54.0% identity in 161 aa overlap

```

```

pep2      10      20      30      40      50      59
EEVPSRLDIRVGKIIIVKEHPDADSLYVEKIDVGEAEPTTVVSGLVQFVKPELDRLY
W14561      10      20      30      40      50      59
EEVPSRLDIRVGKIIIVKEHPDADSLYVEKIDVGEAEPTTVVSGLVQFVKPELDRLY

```

```

pep2      60      70      80      90      100      110      119
VVLNLKPKQMRGVSOGMLCASISGIRNQEPLDPDPASAPGEHVFVYKGGQDDE
W14561      60      70      80      90      100      110      119
VVLNLKPKQMRGVSOGMLCASISGIRNQEPLDPDPASAPGEHVFVYKGGQDDE

```

```

pep2      120      130      140      150      160
LKPKKVFEKLQADFKIISSECIAGWKQTNMTKLSISCSKLGKNGIS
W14561      120      130      140      150      160
LKPKKVFEKLQADFKIISSECIAGWKQTNMTKLSISCSKLGKNGIS

```

```

pep2
GENESQ:R72577
ID R72577 standard; Protein: 310 AA.
AC R72577;
DT 29-SEP-1995 (first entry)
DE Mouse EMAP11.
KW EMAP11; endothelial monocyte activating polypeptide II; chemotaxis;
KW inflammation; tissue factor; tumor; cancer; therapy; metha;

```

```

SCORES      Init1: 398      Initn: 543      Opt: 582 z-score: 855.4 E(): 1.2e-40
Smith-Waterman score: 582; 54.0% identity in 161 aa overlap

```

```

pep2      10      20      30
EEVPSRLDIRVGKIIIVKEHPDADSLYVEKIDVGEAEPTTVVSGLVQFVKPELDRLY
R72577      10      20      30
EEVPSRLDIRVGKIIIVKEHPDADSLYVEKIDVGEAEPTTVVSGLVQFVKPELDRLY

```

```

pep2      40      50      60      70      80      90
EAPRTVVSGLVQFVKPELDRLYVVLNLKPKQMRGVSOGMLCASISGIRNQEPL
R72577      40      50      60      70      80      90
EAPRTVVSGLVQFVKPELDRLYVVLNLKPKQMRGVSOGMLCASISGIRNQEPL

```

```

pep2      100      110      120      130      140      150
DPPASAPGEHVFVYKGGQDDELPKPKVFEKLQADFKIISSECIAGWKQTNMTKLS
R72577      100      110      120      130      140      150
DPPASAPGEHVFVYKGGQDDELPKPKVFEKLQADFKIISSECIAGWKQTNMTKLS

```

```

pep2      160      170      180      190      200      210      220      230
SISCSKLGKNGIS
R72577      160      170      180      190      200      210      220      230
SISCSKLGKNGIS

```

```

pep2      240      250      260      270      280      290
APPNQSVPDRIITFDLFP--GEPDKELNPKKIMEDQIDPLHNAECVATYKKAPEVK-G
R72577      240      250      260      270      280      290
APPNQSVPDRIITFDLFP--GEPDKELNPKKIMEDQIDPLHNAECVATYKKAPEVK-G

```

```

pep2      300      310
KGVCAQTMANSGIK
R72577      300      310
KGVCAQTMANSGIK
GENESQ:W20051
ID W20051 standard; Protein: 657 AA.
AC W20051;
DT 10-SEP-1997 (first entry)

```

DE Methionyl-tRNA synthetase from *Staph. aureus*.
 KW tRNA synthetase; *Bacillus subtilis*; mets; immunological response;
 KW antibody; bacterial infection; adherence; damaged tissue; . . .

SCORES Init1: 101 Initn: 101 Opt: 176 z-score: 250.6 E(): 3.3e-07
Smith-Waterman score: 176; 32.3% Identity In 96 aa overlap

pep2 10 20 30
EEVIPSRLDIRVGKIITVEKHPDADSLYEKIDVGE

```

pep2      AERTVSGIVSGVPPREELQDLVYVLCUKKQKRGVSESGMLTASIGSINRQYE-PL
          . : : : : : : : : : : : : : : : : : : : : : : : :
          : : : : : : : : : : : : : : : : : : : : : : : :
W20051    SEQRQIVSGIAKRYTPDDIIGKRAVWTLKPKRMKGKSGMILLSEKGVTLIVSLPS
          590      600      610      620      630      640

```

pep2 100 110 120 130 140 150
DPPGSAIPGEHYFVKYKGPDEELPKPKKVFKEKLAQDFKISBECLAQNKQNFMTKLG
| |::
W20051 AIPNGAVIK
650

pep2
GENESEQP:W20570

ID	W20570 standard; Protein; 436 AA.
AC	W20570;
DT	17-JUL-1997 (first entry)
DE	H. pylori cytoplasmic protein 785437.aa.
KW	Cytoplasmic; vaccine; prevention; treatment; infection; identification; binding compound; bacterium; life cycle; activator; bacteria; inhibitor; . .

SCORES InIt1: 136 InItN: 136 Opt: 157 z-score: 226.2 E(): 7.4e-06
Smith-Waterman score: 157, 34.7% identity in 75 aa overlap

```

pep2      10      20      30
          EEVIPSRLDINVGKIITVEKHPDADSLIYEKKIDVGE
W20570    ::::| | :: | | ::| |
          EEALAEAKEREKAPPTQENVISDEBKKEVILKQAQRIEKNLIRKVDLGE
          10      20      30

```

pdp2
AEPRTVSGSLGVQVPKEELQDLVVYLCLNAKQKKRGVESGMILCASIGGINRQVEPLT
:
: : : : :
: : : : :
NRIQLIIISGIALDYEESLVGGMCVYANLKRAKLMGMSMGMLAVRDNDNLALLSPTRP
370 380 390 400 410 420

W20570

pep2 100 110 120 130 140 150
P P A G S A N A G E H V F V K G E K G Q P D E L K P K K Y E K I Q A P F K I S E C I A Q M Q T N F M K L G S
W20570 430
E K I N G S L I S

pep2
GENESEQP:W21896

ID	W21896	standard; Protein; 800 AA.
AC	W21896;	
DT	10-SEP-1997	(first entry)
DE	phenylalanyl tRNA synthetase beta subunit.	
KM	tRNA synthetase; <i>Bacillus subtilis</i> ; immunological response; antibody; bacterial infection; adherence; damaged tissue; wound healing; skin; . .	

SCORES Init1: 125 Initn: 147 Opt: 160 z-score: 225.3 E(): 8.4e-06
Smith-Waterman score: 161, 31.7% identity in 142 aa overlap

pep2
W21896

EEVYPSRLDIR--VGKIITVEKHPDADSL
:::| ||: ||: | |||||: |
ISNEMLKVEYTIIDSVSDLAERITRTGIEVDLIDYTKDIKNLVGVFKSEKHPDADKLL

[illegible]

pep2 90 100 110 120 130
ASTIE-GINRVEDLDPAC-----SAGEHVEYVKGEGKQGPDE-ILTRKKVFEKLQA
: |||::: ||::: |||::: |||::: |||:::
w21696 SLAEIGISSNYIKFSESSIVYFSESQVPGIDALQALYLDQVNEFVLTRNRDALSMIGK
130 140 150 160 170 180

DFKISECIAQWKQTNMTKLGISCKSLKGNIS

pep2
SENSESEQ : W20999

ID	W20999	sta
AC	W20999;	

ID	W20999	standard; protein; 662 AA.
AC	W20999;	
DT	21-JUL-1997	(first entry)
DE	H. pylori cytoplasmic protein, hp41452orf1.	
KW	Cytoplasmic; vaccine; prevention; treatment; infection; identification; binding compound; bacterium; life cycle; activator; bacteria; inhibitor; . .	

SCORES Initl: 136 Initn: 136 Opt: 157 z-score: 222.5 E(): 1.2e-05
Smith-Waterman score: 157, 34.7% identity in 75 aa overlap

pep2
W20999

	10	20	30
	EVIPSRDRIIVGKRIIVPEKHPDADSLYEKLVIDGE		
	EEALAEAKKKKKERAPPTQENYI	IEDPFKVEKIVGLIKAEQIESNNLLKRLDGE	
340		560	570
		580	590
540			350

pep2
W20999

	40	50	60	70	80	90
AEPPTVSGLVGCVPEEELQDRLVAVLNLKPKMGREGSOGMLCASIGIRNVEPLD						
: : : : :	:	:	:	:	:	:
NRLRQITSSGIALDYEEESLVGMGVYANLKPALIMGEMSGMILRAYRDNDNLALSPPRR						
600	610	620	630	640	650	

pep2 100 110 120 130 140 150
PPASAGGEHVFVKYEGKQPDDELKPKKVFYFKLQADFISERCIQAQWQNTFMKLGSS

w20999 EKINGSLIS
660

pep2
GENESEQP:R13875

ID R13875 standard; Protein; 248 AA.
 AC R13875:
 DT 11-OCT-1991 (first entry)
 DE Murine Cytotoxic Cell Protease-2.
 KW mouse; CCP2 inhibitor; cytotoxic T-lymphocytes; ss
 OS Mus musculus. . . .

```

SCORES      Init1:   53  Initn:   53  Opt:   101 z-score: 148.8 E():   0.15
Smith-Waterman score: 103;    25.7% identity in 148 aa overlap

```

pep2 20 30 40 50 60 70
 RVGRIITVEKHPDADSLYEKIDVGEAPRIVVSGLVQFVPEKEIDLRLVVLNLCKPKPK
 R13875 KVGKKMFQCGGELVDRKEFWLTIAHCKSSMTVTYGAHNITAKKEETQOIIIVAAITHPDY

```

pep2      80      90      100      110      120
MNGVESOGMLCASIRGINRQ--VEPLDPAGSA---PGEHVFVKYKGEKGPDEELKRP--
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
NPDGRNDIDILKLKLVNNAKTRAVPLNDRRNAHYKPGCECTVAGGKVTGDEGPKTL
110      120      130      140      150      160

pep2      130      140      150      160
-----KKKVE-KLQADF-KISECT--AQMKTNFTKLGS--ISCKSLKGNIS
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
HEKLVQGDQDVCESEFOSSTYRANECYSDSKIKGSEEDSGGPKRAAGSYISG
170      180      190      200      210      220

R13875      QTDGSAQVETRYLSEFVSWIKTKTKHS
230      240

pep2
GENESEQP:R42379
ID      R42379 standard; Protein: 176 AA.
AC      R42379;
DE      19-APR-1994 (first entry)
KW      HaemIn-binding protein produced as fusion protein in pGCH5.
        Haemophilus somnus; immunogenic; haemolysin; lppB; lppC;
        thromboembolic meningoencephalitis; septicæmia; arthritis; . . .

SCORES      Initl: 36 Initn: 36 Opt: 89 z-score: 134.1 E(): 1
Smith-Waterman score: 89; 23.9% identity in 92 aa overlap

pep2      90      100      110      120      130      140
ROVEPLDPAPGAPGEHVFVKYKGEKGPDEELKPKKRYFEKLAQ-----DFKISECIA
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
R--DPYHCPAD-----VLTVIGISTEANGKNIDPKKRISDKRIARMAVDLRLAECVN
50      60      70      80      90

R42379      150      160
OMKQTNFTKLGSISCKSLKGNIS
::
RGNGKNLPOGAFDAVSTFNVGCGKMKSTLEKQANGFTPOLCHQFERMTYAGKKL
100      110      120      130      140      150

pep2
GENESEQP:R42374
ID      R42374 standard; Protein: 178 AA.
AC      R42374;
DE      19-APR-1994 (first entry)
KW      HaemIn-binding protein from ORF5 of plasmid pRAP501.
        Haemophilus somnus; immunogenic; haemolysin; lppB; lppC;
        thromboembolic meningoencephalitis; septicæmia; arthritis; . . .

SCORES      Initl: 36 Initn: 36 Opt: 89 z-score: 134.0 E(): 1
Smith-Waterman score: 89; 23.9% identity in 92 aa overlap

pep2      30      40      50      60      70      80      89
EKIDVGEAPERTVYSGLVQFVKPEELQDLRLVVLNCLKRQKMGVESOGMLCASIEGIN
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
MNFRTKMGTAICSVAAIILVYKANHDELK-ISOQGLDLGNVEGCR
10      20      30      40

R42374      90      100      110      120      130      140
ROVEPLDPAPGAPGEHVFVKYKGEKGPDEELKPKKRYFEKLAQ-----DFKISECIA
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
R--DPYHCPAD-----VLTVIGISTEANGKNIDPKKRISDKRIARMAVDLRLAECVN
50      60      70      80      90

```

```

pep2      150      160
OMKQTNFTKLGSISCKSLKGNIS
::
RGNGKNLPOGAFDAVSTFNVGCGKMKSTLEKQANGFTPOLCHQFERMTYAGKKL
100      110      120      130      140      150

pep2
GENESEQP:R69760
ID      R69760 standard; Protein: 112 AA.
AC      R69760;
DE      05-OCT-1995 (first entry)
        PMON5973 (Met-15-125HIL-3) having an M7 substitution.
        Hybrid: human; murine; interleukin 3; IL-3; deletion; substitution;
        biological activity; side effect profile; leukotriene; release; . . .

SCORES      Initl: 44 Initn: 44 Opt: 81 z-score: 126.4 E(): 2.7
Smith-Waterman score: 81; 28.4% identity in 81 aa overlap

pep2      30      40      50      60      70      80
SLYVEKIDVGEAPERTVYSGLVQFVKPEELQDLRLVVLNCLKRQKRGVE--SQGMLLCAS
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
MNCNMIDEITTHLKQPLPLDLENLNGEDODILMNNLRPNLEAFNRSGGEVDNMS
10      20      30      40      50

pep2      90      100      110      120      130      140
-IEGINROYEPLDPAPGAPGEH-VFVKYKGEKGPDEELKPKKRYFEKLAQDFKISECT
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
AIESILKNLPLCLPLATAPTLRPIHIKGDWNEFRKRTFYKLTLENAQAOO
60      70      80      90      100      110

pep2      150      160
AQMKTNFTKLGSISCKSLKGNIS
GENESEQP:R78335
ID      R78335 standard; Protein: 111 AA.
AC      R78335;
DE      27-JUN-1996 (first entry)
        Human IL-3 mutant for multi-lineage haematopoietic cell prodn.
        Interleukin-3; mutain; colony stimulating factor; CSF; chemotherapy;
        radiation; haematopoiesis; IL-3 antagonist; antigenic; antibody; . . .

SCORES      Initl: 55 Initn: 55 Opt: 80 z-score: 125.0 E(): 3.2
Smith-Waterman score: 80; 23.5% identity in 81 aa overlap

pep2      30      40      50      60      70      80
SLYVEKIDVGEAPERTVYSGLVQFVKPEELQDLRLVVLNCLKRQKRGVE--SQGMLLC
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
MNCNMIDEITTHLKQPLPLDLENLNGEDODILMNNLRPNLEAFNRVKSQNNMS
10      20      30      40      50

pep2      90      100      110      120      130      140
STEGINROYEPLDPAPGAPGEH-VFVKYKGEKGPDEELKPKKRYFEKLAQDFKISECT
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
GIEATLRNLOPCLPLSATAPSRHPIIKAGDWOEFRRKLTFTYKLTLENAQAOO
60      70      80      90      100      110

pep2
GENESEQP:R56050
ID      R56050 standard; Peptide: 111 AA.
AC      R56050;
DE      23-JUN-1995 (first entry)
        Human Interleukin-3 mutain (pMon13350).

```


KW Human Interleukin-3; hIL-3; mutant; mutein.
OS Synthetic. . . .

SCORES Initl: 55 Initn: 55 Opt: 80 z-score: 125.0 E(): 3.2
Smith-Waterman score: 80; 23.5% identity in 81 aa overlap

pep2 SLVEKIDVGEAEPTVVSGLVQFVKKELODRLVVLCNPKPKKVGES--OGMLCA
R56050 NCSNMIDEITTHLKQPLPLDPNNNGEDODILMENNLRPNLEAFNRAVKSQNAS
10 20 30 40 50 60 70 80

pep2 SIEGINROVEPLDPAGSAPGEH-VFVKGEGKQDDEELKPKKVEKQLQADFKESECI
R56050 GIEALRNLOQPLPSATAPSRHPITIKAGDWOEFRRKLTFTYLTLENQAQO
60 70 80 90 100 110

pep2 AOWKOTNFMTKLGSISCKSLKGNIS
GENESEQP: P80381

ID P80381 standard; protein; 132 AA.
AC P80381;
DT 17-NOV-1990 (first entry)
DE Interleukin-3 variant; mutein;
KW Interleukin-3 variant; mutein;
KW colony stimulating factor activity. . . .

SCORES Initl: 44 Initn: 44 Opt: 81 z-score: 124.9 E(): 3.3
Smith-Waterman score: 81; 23.5% identity in 81 aa overlap

pep2 SLVEKIDVGEAEPTVVSGLVQFVKKELODRLVVLCNPKPKKVGES--OGMLCA
P80381 WVNCNMIDEITTHLKQPLPLDPNNNGEDODILMENNLRPNLEAFNRAVKSQNAS
20 30 40 50 60 70 80

pep2 SIEGINROVEPLDPAGSAPGEH-VFVKGEGKQDDEELKPKKVEKQLQADFKESECI
P80381 AIESILKLLPCLPLATAPTRHPHIDGDMNERRKLTFTYLTLENQAQOQTLSLAI
80 90 100 110 120 130

pep2 AOWKOTNFMTKLGSISCKSLKGNIS
P80381 F

pep2
GENESEQP: P80383

ID P80383 standard; protein; 133 AA.
AC P80383;
DT 18-NOV-1990 (first entry)
DE Interleukin-3 variant;
KW Interleukin-3 variant; mutein;
KW colony stimulating factor activity. . . .

SCORES Initl: 44 Initn: 44 Opt: 81 z-score: 124.8 E(): 3.3
Smith-Waterman score: 81; 23.5% identity in 81 aa overlap

pep2 SLVEKIDVGEAEPTVVSGLVQFVKKELODRLVVLCNPKPKKVGES--OGMLCA
P80383 WVNCNMIDEITTHLKQPLPLDPNNNGEDODILMENNLRPNLEAFNRAVKSQNAS
20 30 40 50 60 70 80

pep2 SIEGINROVEPLDPAGSAPGEH-VFVKGEGKQDDEELKPKKVEKQLQADFKESECI
P80383 AIESILKLLPCLPLATAPTRHPHIDGDMNERRKLTFTYLTLENQAQOQTLSLAI
80 90 100 110 120 130

pep2 AOWKOTNFMTKLGSISCKSLKGNIS
P80383 F

pep2
GENESEQP: R56095

ID R56095 standard; peptide; 113 AA.
AC R56095;
DT 26-JUN-1995 (first entry)
DE Human Interleukin-3 mutein (pMon13408).
KW Human Interleukin-3; hIL-3; mutant; mutein.
OS Synthetic. . . .

SCORES Initl: 54 Initn: 54 Opt: 80 z-score: 124.8 E(): 3.3
Smith-Waterman score: 80; 28.8% identity in 80 aa overlap

pep2 SLVEKIDVGEAEPTVVSGLVQFVKKELODRLVVLCNPKPKKVGESOGMLCAS
R56095 MANCSIMIDEITTHLKRPNPPLDPNNLNSEDMIDILIRNRTPLLAFFVRAVKHLENAS
10 20 30 40 50 60 70 80

pep2 -IEGINROVEPLDPAGSAPGEH-VFVKGEGKQDDEELKPKKVEKQLQADFKESECI
R56095 GIEALRNLOQPLPSATAPSRHPITIKAGDWOEFRRKLTFTYLTLENQAQO
70 80 90 100 110

pep2 AOWKOTNFMTKLGSISCKSLKGNIS
pep2

pep2
GENESEQP: R56105

ID R56105 standard; peptide; 113 AA.
AC R56105;
DT 26-JUN-1995 (first entry)
DE Human Interleukin-3 mutein (pMon13380).
KW Human Interleukin-3; hIL-3; mutant; mutein.
OS Synthetic. . . .

SCORES Initl: 54 Initn: 54 Opt: 79 z-score: 123.4 E(): 4
Smith-Waterman score: 79; 28.8% identity in 80 aa overlap

pep2 SLVEKIDVGEAEPTVVSGLVQFVKKELODRLVVLCNPKPKKVGESOGMLCAS
R56105 MANCSIMIDEITTHLKRPNPPLDPNNLNSEDMIDILIRNRTPLLAFFVRAVKHLENAS
10 20 30 40 50 60 70 80

pep2 -IEGINROVEPLDPAGSAPGEH-VFVKGEGKQDDEELKPKKVEKQLQADFKESECI
R56105 GIEALRNLOQPLPSATAPSRHPITIKAGDWOEFRRKLTFTYLTLENQAQO
70 80 90 100 110

pep2 AOWKOTNFMTKLGSISCKSLKGNIS
pep2

pep2
GENESEQP: R56098

pep2
GENESEOP:R56091
ID R56091 standard: peptide; 113 AA.
AC R56091;
DT 26-JUN-1995 (first entry)
DE Human interleukin-3 mutein (pMon13440).
KW Human interleukin-3; hIL-3; mutant; mutein.
OS Synthetic. . . .

SCORES Initl: 54 Infn: 54 Opt: 77 z-score: 120.4 E(): 5.8
Smith-Waterman score: 77; 28.8% identity in 80 aa overlap

pep2
SLYVEKIDVGEAEPRTVVSGLVGFVPEKELODRLVVLCNLK-PQKMGVESQGMILCAS
30 40 50 60 70 80
MANCSIMIDEIHHLKRPNPDLDPNNLNSEDMILMERNLRTPNLLAFVRAVHLENAS
10 20 30 40 50 60
pep2
-IEGINQVEPLDPAGSAPGEH-VFVKGYEKGQDEELPKKKVFEKLQADFRISECI
90 100 110 120 130 140
||:|:::| |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
GIEAILNLQPCLPSTATAAPSRHRPIIITKAGDMQEFREKLFYLVTLQAOEQO
70 80 90 100 110

pep2
AQMKTNFMTRKLSISCKSLKGNIS

pep2
GENESEOP:R78352

ID R78352 standard: protein; 113 AA.
AC R78352;
DT 27-JUN-1996 (first entry)
DE Human IL-3 mutant for multi-lineage haematopoietic cell prodn.
KW Interleukin-3; mutein; colony stimulating factor; CSF; chemotherapy;
radiation; haematopoiesis; IL-3 antagonist; antigenic; antibody; . . .

SCORES Initl: 54 Infn: 54 Opt: 77 z-score: 120.4 E(): 5.8
Smith-Waterman score: 77; 28.8% identity in 80 aa overlap

pep2
SLYVEKIDVGEAEPRTVVSGLVGFVPEKELODRLVVLCNLK-PQKMGVESQGMILCAS
30 40 50 60 70 80
MANCSIMIDEIHHLKRPNPDLDPNNLNSEDMILMERNLRTPNLLAFVRAVHLENAS
10 20 30 40 50 60
pep2
-IEGINQVEPLDPAGSAPGEH-VFVKGYEKGQDEELPKKKVFEKLQADFRISECI
90 100 110 120 130 140
||:|:::| |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
GIEAILNLQPCLPSTATAAPSRHRPIIITKAGDMQEFREKLFYLVTLQAOEQO
70 80 90 100 110

pep2
AQMKTNFMTRKLSISCKSLKGNIS

pep2
GENESEOP:R56137

ID R56137 standard: peptide; 113 AA.
AC R56137;
DT 28-JUN-1995 (first entry)
DE Human interleukin-3 mutein (pMon13440).
KW Human interleukin-3; hIL-3; mutant; mutein.
OS Synthetic. . . .

SCORES Initl: 54 Infn: 54 Opt: 77 z-score: 120.4 E(): 5.8
Smith-Waterman score: 77; 28.8% identity in 80 aa overlap

30 40 50 60 70 80

pep2
SLYVEKIDVGEAEPRTVVSGLVGFVPEKELODRLVVLCNLK-PQKMGVESQGMILCAS
30 40 50 60 70 80
MANCSIMIDEIHHLKRPNPDLDPNNLNSEDMILMERNLRTPNLLAFVRAVHLENAS
10 20 30 40 50 60
pep2
-IEGINQVEPLDPAGSAPGEH-VFVKGYEKGQDEELPKKKVFEKLQADFRISECI
90 100 110 120 130 140
||:|:::| |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
GIEAILNLQPCLPSTATAAPSRHRPIIITKAGDMQEFREKLFYLVTLQAOEQO
70 80 90 100 110

pep2
AQMKTNFMTRKLSISCKSLKGNIS

pep2
GENESEOP:R78368

ID R78368 standard: protein; 113 AA.
AC R78368;
DT 27-JUN-1996 (first entry)
DE Human IL-3 mutant for multi-lineage haematopoietic cell prodn.
KW Interleukin-3; mutein; colony stimulating factor; CSF; chemotherapy;
radiation; haematopoiesis; IL-3 antagonist; antigenic; antibody; . . .

SCORES Initl: 54 Infn: 54 Opt: 77 z-score: 120.4 E(): 5.8
Smith-Waterman score: 77; 28.8% identity in 80 aa overlap

pep2
SLYVEKIDVGEAEPRTVVSGLVGFVPEKELODRLVVLCNLK-PQKMGVESQGMILCAS
30 40 50 60 70 80
MANCSIMIDEIHHLKRPNPDLDPNNLNSEDMILMERNLRTPNLLAFVRAVHLENAS
10 20 30 40 50 60
pep2
-IEGINQVEPLDPAGSAPGEH-VFVKGYEKGQDEELPKKKVFEKLQADFRISECI
90 100 110 120 130 140
||:|:::| |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
GIEAILNLQPCLPSTATAAPSRHRPIIITKAGDMQEFREKLFYLVTLQAOEQO
70 80 90 100 110

pep2
AQMKTNFMTRKLSISCKSLKGNIS

pep2
GENESEOP:R56120

ID R56120 standard: peptide; 113 AA.
AC R56120;
DT 26-JUN-1995 (first entry)
DE Human interleukin-3 mutein (pMon13383).
KW Human interleukin-3; hIL-3; mutant; mutein.
OS Synthetic. . . .

SCORES Initl: 54 Infn: 54 Opt: 77 z-score: 120.4 E(): 5.8
Smith-Waterman score: 77; 31.3% identity in 83 aa overlap

pep2
DSLYVEKIDVGEAEPRTVVSGLVGFVPEKELODRLVVLC--NLK-PQKMGVESQGMIL
30 40 50 60 70 80
MANCSIMIDEIHHLKRPNPDLDPNNLNSEDMILMERNLRTPNLLAFVRAVHLENAS
10 20 30 40 50
pep2
CAS-IEGINQVEPLDPAGSAPGEH-VFVKGYEKGQDEELPKKKVFEKLQADFRISE
90 100 110 120 130 139
||:|:::| |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
NASGIEAILNLQPCLPSTATAAPSRHRPIIITKAGDMQEFREKLFYLVTLQAOEQO
60 70 80 90 100 110

pep2
ECIAQMKTNFMTRKLSISCKSLKGNIS

pep2

OS Synthetic.

DT 22-JUN-1995 (first entry)

DE Human Interleukin-3 mutein (pMon13289).
 KW Human Interleukin-3; hIL-3; mutant; mutein.
 OS Synthetic. . . .

SCORES Initl: 54 Initn: 54 Opt: 76 z-score: 118.9 E(): 7
 Smith-Waterman score: 76; 28.8% identity in 80 aa overlap

pep2 SLVEKIDVGEAEPRTVVSGLVGFPEKELQDLVYVLCNLK-PQKMGVESQGMILCAS
 30 40 50 60 70 80
 R56068 MANCSIMDELIHHKRPNPPLDDNNNSDEVDILMRNLRTPLLAFAVAVKLENAS
 10 20 30 40 50 60

pep2 -IEGINRQVEPLDPAGSAPGEH-VFVKGYEKGQDPEELPKKKVFEKLQADFRISEECI
 90 100 110 120 130 140
 R56068 GIEALIRNLQPCLPASATAPSRHRPIITIKAGDMQEFREKLTFTYVLTLEQAQEQ
 70 80 90 100 110

pep2 AQMKTNFMTRKLSISCKSLKGNIS

pep2
 GENESQ:R56092

ID R56092 standard; peptide; 113 AA.

AC R56092;

DT 26-JUN-1995 (first entry)

DE Human Interleukin-3 mutein (pMon13407).

KW Human Interleukin-3; hIL-3; mutant; mutein.

OS Synthetic. . . .

SCORES Initl: 54 Initn: 54 Opt: 76 z-score: 118.9 E(): 7
 Smith-Waterman score: 76; 28.8% identity in 80 aa overlap

pep2 SLVEKIDVGEAEPRTVVSGLVGFPEKELQDLVYVLCNLK-PQKMGVESQGMILCAS
 30 40 50 60 70 80
 R56092 MANCSIMDELIHHKRPNPPLDDNNNSDEVDILMRNLRTPLLAFAVAVKLENAS
 10 20 30 40 50 60

pep2 -IEGINRQVEPLDPAGSAPGEH-VFVKGYEKGQDPEELPKKKVFEKLQADFRISEECI
 90 100 110 120 130 140
 R56092 GIEALIRNLQPCLPASATAPSRHRPIITIKAGDMQEFREKLTFTYVLTLEQAQEQ
 70 80 90 100 110

pep2 AQMKTNFMTRKLSISCKSLKGNIS

pep2
 GENESQ:R56101

ID R56101 standard; peptide; 113 AA.

AC R56101;

DT 26-JUN-1995 (first entry)

DE Human Interleukin-3 mutein (pMon13425).

KW Human Interleukin-3; hIL-3; mutant; mutein.

OS Synthetic. . . .

SCORES Initl: 54 Initn: 54 Opt: 76 z-score: 118.9 E(): 7
 Smith-Waterman score: 76; 28.8% identity in 80 aa overlap

pep2 SLVEKIDVGEAEPRTVVSGLVGFPEKELQDLVYVLCNLK-PQKMGVESQGMILCAS
 30 40 50 60 70 80
 R56101 MANCSIMDELIHHKRPNPPLDDNNNSDEVDILMRNLRTPLLAFAVAVKLENAS
 10 20 30 40 50 60

pep2 -IEGINRQVEPLDPAGSAPGEH-VFVKGYEKGQDPEELPKKKVFEKLQADFRISEECI
 90 100 110 120 130 140
 R56101 GIEALIRNLQPCLPASATAPSRHRPIITIKAGDMQEFREKLTFTYVLTLEQAQEQ
 70 80 90 100 110

pep2 AQMKTNFMTRKLSISCKSLKGNIS

! CPU time used:
 ! Database scan: 0:01:37.3
 ! Post-scan processing: 0:00:02.9
 ! Total CPU time: 0:01:40.2
 ! Output file: pep2.fasta

! IAA SEQUENCE 1.0

ID W06596 standard; protein; 168 AA.

AC W06596;

DT 04-JUL-1997 (first entry)

DE Human endothelial-monocyte activating polypeptide III.

KW Endothelial-monocyte activating polypeptide; EMAP III; human;

KW neoplasia.

OS Synthetic.

PN W09640719-A1.

PD 19-DEC-1996.

PE 07-JUN-1995; U07328.

PR 07-JUN-1995; W0-U07328.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Coleman TA, Olsen HS, Rosen CA;

DR WPI: 97-052219/05.

DR N-PSDS; T49312.

PT New isolated endothelial-monocyte activating polypeptide III - used

PT to develop prods. for the diagnosis and treatment of conditions

PT involving EMAP III, e.g. neoplasia

PS Claim 4: Page 38; 50pp; English.

CC The present sequence represents the 168 amino acid sequence for the novel

CC polypeptide endothelial-monocyte activating polypeptide III (EMAP III).

CC The product can be used for the diagnosis and treatment of conditions

CC involving abnormal expression of EMAP III. In particularly EMAP III can

CC be used to regress neoplasia such as in tumours and cancers.

CC EMAP III shows a high degree of homology to EMAP II with 60%

CC identity and 75% similarity over a 150 amino acid stretch.

SQ Sequence 168 AA;

SQ 7 A; 6 R; 4 N; 8 D; 0 B; 4 C; 9 Q; 18 E; 0 Z; 13 G; 2 H;

SQ 11 I; 14 L; 18 K; 3 M; 5 F; 11 P; 11 S; 4 T; 1 W; 2 Y; 17 V;

W06596 Length: 168 April 28, 1998 12:48 Type: P Check: 3813 ..

1 EEVIPSRLDI RVGKITVER HPDADSLYE KIDVGEAEPR TVVSGLVGFV

51 PREELQDLRV VVLCNLKPKQ MGVESQGM LCAISIEGINR QVEPLDPAG

101 SAGEHVFVK GYKQGPDEE LKPKKVFKE LQADFRISEE CIQMKTNFM

151 MTRKLSISCK SLKGNIS

! IAA SEQUENCE 1.0

ID W14561 standard; protein; 166 AA.

AC W14561;

DT 01-DEC-1997 (first entry)

DE Endothelial monocyte activating polypeptide II.

KW Endothelial monocyte activating polypeptide; blood vessel;

KW EMAP-II; inhibitor; endothelial cell formation; blood vessel;

KW retinopathy; tumour; subcutaneous; intraperitoneal; intravenous;

KW intracranial; intratumoural; carcinoma; glioblastoma;

KW positive pressure microinfusion; aortic cell; angiogenesis;

KW eye disease; diabetes; sickle cell anaemia; prematurity;

KW age-related macular degeneration.

OS Mus musculus.

PH Key location/Qualifiers

FT Misc_difference 1

FT /label= Ser, Met, Gly

FT Misc_difference 15

FT /label= Cys, Arg

PN W09710841-A1.
 PD 27-MAR-1997.
 PR 18-SEP-1996; U15007.
 PR 18-SEP-1995; US-003898.
 PA (UYCO) UNIV COLUMBIA NEW YORK.
 PI Schwartz M, Stern D;
 DR WPI: 97-202617/18.
 PT Treatment of tumours with endothelial monocyte activating peptide II
 PT - also used to inhibit growth of endothelial cells and formation of
 PT blood vessels, e.g. in retinopathy
 PS Claim 13; Page 73; 104pp; English.
 CC A method has been developed for the treatment of tumours. The method
 CC involves subcutaneous, intraperitoneal, intravenous, intracranial or
 CC intratumoural administration of an endothelial monocyte activating
 CC polypeptide II (EMAP-II), or a EMAP-II-derived polypeptide. The present
 CC sequence represents murine EMAP-II which can be used to produce the
 CC EMAP-II-derived polypeptide as long as it contains at least ninety
 CC percent homology to the present sequence. The method is especially used
 CC to treat carcinoma (administered intraperitoneally) or glioblastoma
 CC (administered intratumourally by positive pressure microinfusion).
 CC Inhibition of endothelial cell growth is particularly applied to aortic
 CC cells and inhibition of angiogenesis to treatment of eye diseases
 CC associated with excessive blood vessel formation, especially retinopathy
 CC (diabetic, related to sickle cell anaemia, retinopathy of prematurity or
 CC age-related macular degeneration). Intraperitoneal administration allows
 CC tumours too small (not over 2 mm in diameter) for intratumoural
 CC injection to be treated before they grow larger.
 SO Sequence 166 AA;
 SO 15 A; 7 R; 7 N; 9 D; 0 B; 4 C; 6 Q; 11 E; 0 Z; 11 G; 3 H;
 SO 8 I; 12 L; 14 K; 6 M; 3 F; 14 P; 8 S; 6 T; 1 W; 2 Y; 17 V;
 SO 2 Others;

W14561 Length: 166 April 28, 1998 12:48 Type: P Check: 5793 ..

1 NKPIASRLD LRIGNITAK KHPDADSLY EEDVNGEAP RTVSGLYNH
 51 VPLEQNMNM VLLCNLKP KMGVLSQAM VMCASSPEV ELIAPNGSV
 101 PGDRITFDNF PGEPDKELNP KKKIMEQIOP DLHTNACVA TYKGAPFEVK
 151 GKGVRAQTM ANSGIK

!!AA_SEQUENCE 1.0
 ID R72577 standard; Protein: 310 AA.
 AC R72577;
 DT 29-SEP-1995 (first entry)
 DE Mouse EMAPII.
 KW EMAPII; endothelial monocyte activating polypeptide II; Chemotaxis;
 KW inflammation; tissue factor; tumor; cancer; therapy; metha;
 KW sarcoma.
 OS Mus musculus.
 PN W09509180-A.
 PD 06-APR-1995.
 PR 29-SEP-1994; U11085.
 PR 29-SEP-1993; US-129456.
 PA (UYCO) UNIV COLUMBIA NEW YORK.
 PI Claus M, Kao J, Kayton M, Libutti SK, Stern DM;
 DR WPI: 95-147389/19.
 DR N-PSDB; Q86718.
 PT New endothelial monocyte activating polypeptide II - induces
 PT chemotaxis, inflammation and tissue factor, useful for treating
 PT tumours, also related antibodies, DNA and active fragments
 PS Disclosure; Fig.4; 180pp; English.
 CC A mouse metha sarcoma cDNA library was screened with a probe
 CC based on the N-terminal sequence of mouse EMAPII. Overlapping
 CC clones were combined to obtain a contiguous full-length sequence
 CC (given in Q86718) encoding a 33 kDa protein (R72577).
 CC Recombinant EMAPII was expressed in E. coli.
 SO Sequence 310 AA;
 SO 32 A; 12 R; 11 N; 12 D; 0 B; 6 C; 18 Q; 32 E; 0 Z; 15 G; 4 H;
 SO 15 I; 25 L; 36 K; 8 M; 3 F; 17 P; 17 S; 16 T; 1 W; 3 Y; 27 V;

R72577 Length: 310 April 28, 1998 12:48 Type: P Check: 1208 ..

1 MATNDVILKR LEQGAEDQ IIEYKQOVA LKKEKAIQA TMREKKLRY
 51 ENAKLKEIE ELKEELIAE IHNGVQAVY RLTSLQINC TASSVQSP
 101 SVATTASPAT KEQIKAGEER KYREKTEKKG EKKEKQSSAA ASTDSKPDA
 151 SRLDRICGI VTAKKHDPAD SLVEEVDVG EAAPRTVSG LVNHVPLEQM
 201 QNRMYVILCN LKPAKMGVL SQAMWCASS PEKVELIAP NGSPGDRIT
 251 FDAFGEEDK ELNPKKIWE QIOPDLHTNA ECVATYKGP FEVKGVCYR
 301 AQTMANSQIK

!!AA_SEQUENCE 1.0
 ID W20051 standard; Protein: 657 AA.
 AC W20051;
 DT 10-SEP-1997 (first entry)
 DE Methionyl-tRNA synthetase from Staph. aureus.
 DE tRNA synthetase; Bacillus subtilis; met; immunological response;
 KW antibody; bacterial infection; adherence; damaged tissue;
 KW wound healing; skin; protection; vaccine.
 OS Staphylococcus aureus.
 SO EP-785268-A1.
 PD 23-JUL-1997.
 PR 17-JAN-1997; 300317.
 PR 19-JAN-1996; GB-001095.
 PR 27-JUL-1996; GB-015845.
 PA (SMIK) SMITHKLINE BEECHAM PLC.
 PI Hodgson JE, Lawlor EJ;
 DR WPI: 97-365943/34.
 DR N-PSDB; T71309.
 PT DNA encoding methionyl-tRNA synthetase from Staphylococcus aureus
 PT WCUH 29 - useful for protection against bacterial infections
 PS Claim 1; Pages 19-20; 31pp; English.
 CC The present sequence is a novel methionyl tRNA synthetase
 CC protein from staphylococcus aureus strain WCUH29 (NCIB 40771),
 CC which is related by amino acid sequence homology to Bacillus
 CC subtilis tRNA synthetase encoded by the metS gene. The enzyme
 CC catalyses the aminoacylation of tRNA-Met, by a two step mechanism.
 CC The first step involves formation of a stable enzyme:methionyl
 CC adenylyate complex resulting from the specific binding and reaction
 CC of ATP and L-methionine. Subsequently, the 3' terminal adenosine of
 CC enzyme-bound tRNA-Met reacts with the aminoacyl adenylyate, leading
 CC to the esterification of the tRNA and release of AMP. Vectors
 CC comprising the DNA (or polynucleotides having at least 70 %
 CC identity to it) can be used for the recombinant production of the
 CC enzyme. The enzyme or its related DNA (through gene therapy) is
 CC used to induce an immunological response in a mammal to generate
 CC antibodies to protect against disease. The antibodies protect
 CC against invasion of bacteria, e.g. by blocking adherence of
 CC bacteria to damaged tissue, including wounds in skin or connective
 CC tissue caused by mechanical, chemical or thermal damage or by
 CC implantation of in-dwelling devices, or wounds in the mucous
 CC membranes.
 SO Sequence 657 AA;
 SO 40 A; 20 R; 27 N; 43 D; 0 B; 0 C; 24 Q; 55 E; 0 Z; 36 G; 15 H;
 SO 51 I; 57 L; 51 K; 20 M; 26 F; 35 P; 41 S; 31 T; 9 W; 32 Y; 44 V;

W20051 Length: 657 April 28, 1998 12:48 Type: P Check: 1909 ..

1 MAKEPFYITP PIYPSGNLH IGHAYSTVAG DVIARKKMQ GYDVRILTGT
 51 DEHGKIOEK AQKAGKTEIE YLDEMIAGIK QUMAKLEISN DDFIITTEER
 101 HKHVEQVYE RLKQGDYIL GEYEGWYSP DETYTTESQL VDPQYENKGI
 151 ICGKSPDSGH EVELYKEESY FFINISKYTR LLEFDQNDP FIQPPSKRNE
 201 MINNFIKPOL ADLAVSRTSF NMGVHPSNP KHVVYVWIDA LVNYISALGY
 251 LSDDSLFLNK YMPADIHMA KEIVRFHSII WILMALDUL PLPKVFAHG

301 WLMKDGKMS KSKGNVDPN ILIDRYGLDA TRYLMRELPG FSGDGVETPE
 351 AFVERTNEDL ANDLGNLVNR TISMVKNYFD GELPAYQGPL HELDEMEAM
 401 ALEEVKASTE SMESLOSVA LSTVWKFISS TKNYIDETTP WYLAKDSOK
 451 DMLGNVAHL VENIRYRAVL LRPELTNAPK EIEFQNLNIN PQFMEFSSLE
 501 QYGVLTESIM VTGQPKPIFP RLDSEALIAV IKESMQPPAT EEEKETPSK
 551 PQDIKQEDK VEIKATIID AEVKKSDKL LKIQVDLDE QROIVSGIAK
 601 FYTPDILIGK KVAVTNLKP AKLMQKSEK MILSAEKDGV LTVLSLPSAI
 651 PNGAVIK

11AA_SEQUENCE 1.0
 ID W20570 standard; Protein; 436 AA.
 AC W20570;
 DT 17-JUL-1997 (first entry)
 DE H. pylori cytoplasmic protein 785437.aa.
 KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
 KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
 KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
 OS Helicobacter pylori.
 PN W09640893-A1.
 PD 19-DEC-1996.
 PE 06-JUN-1996; U09122.
 PR 07-JUN-1995; US-487032.
 PR 01-APR-1996; US-630405.
 PA (ASTR) ASTRA AB.
 PI Berglindh OJ, Smith D, Mellgaard BJ;
 DR WPI: 97-052306/05.
 DR N-PSDB: T67713.
 PT Helicobacter pylori nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
 PT infection, and to detect Helicobacter
 PS Claim 61, Page 724-725, 1481pp; English.
 CC This sequence is a H. pylori cytoplasmic protein involved in
 CC mRNA translation and ribosome biogenesis.
 CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds,
 CC useful as potential H. pylori life cycle activators or inhibitors.
 CC The genomic sequence of H. pylori (ATCC 55679) was determined from
 CC overlapping contigs generated by mechanically shearing the bacterial
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
 CC and the predicted coding regions defined by computer evaluation. To
 CC identify likely H. pylori antigens for vaccine development, the amino
 CC acid sequences predicted from various ORF were analysed for significant
 CC homology to other known or exported membrane proteins. Having identified
 CC and determined the sequences of interest, particular regions can be
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide
 CC production, e.g. in E. coli hosts.
 SO Sequence 436 AA;
 SO 35 A; 15 R; 28 N; 17 D; 0 B; 2 C; 12 Q; 42 E; 0 Z; 19 G; 8 H;
 SO 30 I; 61 L; 48 K; 14 M; 18 F; 15 P; 23 S; 8 T; 5 W; 15 Y; 21 V;
 W20570 Length: 436 April 28, 1998 12:48 Type: P Check: 8042 ..

1 MNDKRVVYV WLDALNTAS ALGYLGLDN KMAFERARH IVGKILRPH
 51 AIWPAFLMS LNLPLFKOLC VHGWTIEGV KMSKSLGNVL DAQKLAIEYG
 101 IEELRYFLLR EYFEGGDGF SKKALVERIN ANLNNDLGNL LNRLLGAAKK
 151 YFVNSLSTK ITAYYKKELE KAHQIIDNAN SFVPMQOLK ALEELFIYD
 201 FLNKLIAKEE PAVLHKNNES EKLEALLSLI ANTLQSSFL LYAPMPSAM
 251 KLASAFNVEI TPNNYERFFK AKKLQDMVQ DTEPLFSKIE KIEKIEIKK
 301 IEKIEKEEBA LAEKAERKKK EKAPTOENY ISIEDFKKVE IKVGLIKEAQ

351 RIEKSNKLLR LKVDLGENRL RQIISGIALD YEPESLVGOM VCYVANKLPA
 401 KLMESEBGM IIAVRNDNM ALISPTREKI AGSLIS

11AA_SEQUENCE 1.0
 ID W21896 standard; Protein; 800 AA.
 AC W21896;
 DT 10-SEP-1997 (first entry)
 DE Phenylalanyl tRNA synthetase beta subunit.
 KW tRNA synthetase; Bacillus subtilis; immunological response; antibody;
 KW bacterial infection; adherence; damaged tissue; wound healing; skin;
 KW protection; vaccine.
 OS Staphylococcus aureus.
 PN EP-785262-A1.
 PD 23-JUL-1997.
 PE 17-JAN-1997; 300311.
 PR 19-JAN-1996; GB-001096.
 PR 27-JUL-1996; GB-015845.
 PA (SMIK) SMITHKLINE BEECHAM PLC.
 PI Hodgson JE, Lawlor EJ;
 DR WPI: 97-365937/34.
 DR N-PSDB: T73694.
 PT DNA encoding phenylalanyl-tRNA synthetase alpha subunit from
 PT Staphylococcus aureus WCUH 29 - useful for protection against
 PT bacterial infections
 PS Disclosure: Page 22-24; 41pp; English.
 CC The present sequence is a novel phenylalanyl tRNA synthetase beta subunit
 CC from Staphylococcus aureus strain WCUH29 (NCIMB 40771), which is related
 CC by amino acid sequence homology to Bacillus subtilis phenylalanyl tRNA
 CC synthetase beta subunit. The enzyme catalyses the aminoacylation of
 CC tRNA-Phe, by a two step mechanism. The first step involves formation
 CC of a stable enzyme: phenylalanyl adenylate complex resulting from the
 CC specific binding and reaction of ATP and L-phenylalanine. Subsequently,
 CC the 3' terminal adenosine of enzyme-bound tRNA-Phe reacts with the
 CC aminoacyl adenylate, leading to the esterification of the tRNA and
 CC release of AMP. Vectors comprising the DNA (or polynucleotides having at
 CC least 70% identity to it) can be used for the recombinant production of
 CC the enzyme. The enzyme or its related DNA (through gene therapy) is used
 CC to induce an immunological response in a mammal to generate antibodies to
 CC protect against disease. The antibodies protect against invasion of
 CC bacteria, e.g. by blocking adherence of bacteria to damaged tissue,
 CC including wounds in skin or connective tissue caused by mechanical,
 CC chemical or thermal damage or by implantation of in-dwelling devices,
 CC or wounds in the mucous membranes.
 SO Sequence 800 AA;
 SO 54 A; 38 R; 35 N; 64 D; 0 B; 4 C; 33 Q; 63 E; 0 Z; 52 G; 11 H;
 SO 67 I; 72 L; 42 K; 15 M; 25 F; 28 P; 49 S; 48 T; 3 W; 28 Y; 69 V;
 W21896 Length: 800 April 28, 1998 12:48 Type: P Check: 2312 ..

1 MLINENLKE YVYIDSVSD LAERTIRGI EYVDLIDYTK DINLVVGFV
 51 KSKKEHPDAD KLVNCOVDIG EDEPVQIVCG APNVDAQVY IVAKVGRLP
 101 GGIRIKRAKL RGERSEGMIC SLQETIGISS YIPSPFSGI YVESQVPG
 151 TDAQALYLD DQVMEFDLTP NRADALSMIG TAYEVALYN TKTTRPETTS
 201 NEDELNANDE LVTIENEDK VPIYSARVH DVTEPSPIW MQARLIKAGI
 251 RPINNVDIS NYVLEYGOP LHMEPDQAIG SQQIVRQAN EGEKMTLLD
 301 TERELLTSDI VITNGOTPIA LAGVAGGQFS EYKQTSNIV IEBAINDPVS
 351 IRTSRLNL RSESSRFEK GIATEFVDEA VDRACYLLQT YANGKVLKDR
 401 VSSGELGAFI TPIDITADKI NRTIGFDLSQ NDIYTFNQL GFDTETINDV
 451 ITIVVPSRRK DITIKEDLIE EVARIYGYD IPSTLPVFDK VTSQGLTDRO
 501 YTRVREVL EGAGLDQAIF YSLVSKEDAT AFSMQOROTI DLLMPNSEAH


```

551 ASLSQSLPH LIEVASINVA RKNQVKLFE IGWFFANGE GELPDQVEYL
601 SGILTGIVV NQWQKKEV DFLAKGYD RYSEKLNIEF SYRRADIDL
651 HPGTAEILL ENKVGIGE LHPTLAADND LKRTVFELN FDLAMAVSG
701 YINQPIPRF PGMSRDIALE VDQNIIPADL LSTIHAGGN ILKDTLVEDV
751 YQGELEKKG KSIAIRLNTL DTEETLTDER VSKVQAEIEA ALIDQAVIR

!!A_SEQUENCE 1.0
ID W20999 standard: protein; 662 AA.
W20999;
DT 21-JUL-1997 (first entry)
DE H. pylori cytoplasmic protein, hp4el4522orf11.
KM Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KM binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KM duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
OS Helicobacter pylori.
PN W09640893-A1.
PD 19-DEC-1996.
PF 06-JUN-1996: U09122.
PR 07-JUN-1995: US-487032.
PR 01-APR-1996: US-630405.
PA (ASTR ) ASTRA AB.
PI Berglindh OT, Smith D, Møllgaard BL;
WPI: 97-052306/05.
DR N-PSDB: T68252.
PT Helicobacter pylori nucleic acid sequences and related
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT infection, and to detect Helicobacter
PS Claim 61, Page 1389-1390: 1481pp: English.
CC The present sequence is a H. pylori cytoplasmic protein involved in
CC mRNA translation and ribosome biogenesis.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors.
CC The genomic sequence of H. pylori (ATCC 55679) was determined from
CC overlapping contigs generated by mechanically shearing the bacterial
CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC and the predicted coding regions defined by computer evaluation. To
CC identify likely H. pylori antigens for vaccine development, the amino
CC acid sequences predicted from various ORF were analysed for significant
CC homology to other known or exported membrane proteins. Having identified
CC and determined the sequences of interest, particular regions can be
CC isolated from H. pylori by PCR amplification for recombinant polypeptide
CC production, e.g. in E. coli hosts.
SQ Sequence 662 AA;
SQ 46 A; 21 R; 37 N; 30 D; 0 B; 8 C; 21 Q; 60 E; 0 Z; 30 G; 12 H;
SQ 48 I; 80 L; 66 K; 16 M; 31 F; 24 P; 39 S; 26 T; 7 W; 31 Y; 29 V;
W20999 Length: 662 April 28, 1998 12:48 Type: P Check: 5242 ..

1 GNTIIKQKS LITPIIYV DIPHGHAYT TLIDTLKKY YTLQGEVFF
51 LTGTDEHGOK IEQSARLBNQ SPKAVADSTS ALFNQWDF NLDYDGIIRT
101 TDSHQKCVQ NAFIMFEKG DIYKGTISY YCVSCESCVA VSKVNDTDSK
151 VLPDCLRET TLLEESYFF KLSAYEKPL EYAKNPEAI LPIYRKNQEV
201 SFIEQGLDL SITRTSEWG IPLPKMNDP KHVYVWMDA LNAVASLGY
251 LINGDNKMAH FERARHIVG DILRHAITYW PAFMLSLNLP LFKQLCVHGW
301 WTIEGVMSK SLGNVDAOK LAMEYIEEL RYFLREVPF GODGDEFSKKA
351 LVERINANLN NDGNLNLRL LGMAKKYFN SUKSTRIYAY YPKLELAHQ
401 ILDNANFVP KMLHKALEE LENITDFLAK LIAKEEPVL HKNNESEKLE
451 ALSLIANTL LOSSFLIYAF MPKSAKLAS AFWEITPPNN YERFFKAKKL

```

```

501 QDNVLODTEP LFSKIEKIEK IEKIEKIEK EKGEEALAEK AEKKEKAP
551 PTQENYISIE DFKKVEIKV LKEAQRIEK SNKLLRLKVD LGENRLROI
601 SGILDYEPE SLVGQVCCV ANLKPAKLMG EMSEGMIIAV RDNDAIALIS
651 PTRREKISL IS

!!A_SEQUENCE 1.0
ID R13875 standard: protein; 248 AA.
R13875;
DT 11-OCT-1991 (first entry)
DE Murine Cytotoxic Cell Protease-2.
KM mouse; CCP2 inhibitor; cytotoxic T-lymphocytes; ss.
OS Mus musculus.
PN W09110685-A.
PD 25-JUL-1991.
PF 17-JAN-1991: U00340.
PR 19-JAN-1990: US-467880.
PA (SERA-) SERAGEN INC.
PI Bleackley RC, Lobe CG, Paetkau VH, James MN, Murphy M;
WPI: 91-23789/32.
DR N-PSDB: Q12863.
PT DNA vectors, and inhibitors of cytotoxic cell protease - for
PT treatment of auto-immune diseases e.g. pernicious anaemia,
PT rheumatoid arthritis, allo-graft rejection etc.
PS Claim 6, Fig 10: 62pp: English.
CC The CCP2 coding sequence was isolated from the cytotoxic T-cell
CC line MTL 2.8.2 generated from CBA/J mice. The amino acid sequence
CC of CCP2 was predicted from the cDNA sequence. The structure of the
CC protein was used to design peptides which competitively inhibit the
CC protease. See Q12862-6 and R13254-R13262.
SQ Sequence 248 AA;
SQ 19 A; 13 R; 9 N; 11 D; 0 B; 6 C; 8 Q; 13 E; 0 Z; 20 G; 7 H;
SQ 13 I; 20 L; 21 K; 6 M; 9 F; 16 P; 16 S; 12 T; 2 W; 7 Y; 20 V;
R13875 Length: 248 April 28, 1998 12:48 Type: P Check: 6683 ..

1 MPVLILTL LPLRGAEE ITGNEISPH SRPYMAYEF LKYGKKMFC
51 GGLVRDKFV LTAHCKGSS MTVLGAHNI KAKEEQOII PVAKAIPHPD
101 YNPDRSNDI MLKLVBNK RTRAVRPLN PRRAHVKPG DECVYAGWK
151 VTPGEPKPT LHEVKTIVQK DOVCESQFOS SYNANEICV GDSKINGASF
201 EEDSGPLVC KRAAGIVSY GQTDGAPQV FTRVLSFVSW IKKTMKHS

!!A_SEQUENCE 1.0
ID R42379 standard: protein; 176 AA.
R42379;
DT 19-APR-1994 (first entry)
DE Haemin-binding protein produced as fusion protein in pcgH5.
KM Haemophilus somnus; immunogenic; haemolysin; lppb; lppc;
KM thromboembolic meningencephalitis; septicaemia; arthritis;
KM pneumonia; lktA gene; Pasteurella haemolytica.
OS Haemophilus somnus.
PN W09321323-A.
PD 28-OCT-1993.
PF 05-APR-1993: CA0135.
PR 09-APR-1992: US-865050.
PR 04-JUN-1992: US-893424.
PR 04-JUN-1992: US-893426.
PR 29-MAR-1993: US-038287.
PR 29-MAR-1993: US-038288.
PR 29-MAR-1993: US-038719.
PA (UYSA-) UNIV SASKATCHEWAN.
PI Harland RJ, Pfeiffer CG, Pontarollo RA, Potter AA;
PI Rioux C, Theisen M;
DR WPI: 93-351733/44.
DR N-PSDB: Q51081.
PT Haemophilus somnus immunogenic proteins used in vaccines -
PT selected from haemin-binding protein, haemolysin, lppb and lppc,

```

PT and corresp. DNA
 PS Claim 1; Fig 5; 119pp; English.
 CC The hmb gene encoding the haemin-binding protein was expressed in
 CC E. coli as a fusion to the Pasteurella haemolytica leukotoxin gene
 CC lktA coded for by plasmid pHA352. The hmb gene fragment was taken
 CC from pRAP501 and starts at the codon for the third amino acid residue
 CC of ORF1. The haemin binding protein can be used in vaccines for
 CC preventing or treating H. somnus infections, which cause thromboembolic
 CC meningo-encephalitis, septicemia, arthritis and pneumonia in
 CC vertebrates.
 CC See also R42370-86.
 SQ Sequence 1/6 AA; 8 D; 0 B; 7 C; 11 Q; 9 E; 0 Z; 18 G; 4 H;
 SQ 16 A; 10 R; 10 N; 8 D; 0 B; 7 C; 11 Q; 9 E; 0 Z; 18 G; 4 H;
 SQ 10 I; 14 L; 14 K; 1 M; 7 F; 5 P; 6 S; 7 T; 3 W; 6 Y; 10 V;

R42379 Length: 176 April 28, 1998 12:48 Type: P Check: 5317 ..

1 EFTKWTGAI CSVAIIALV KAHQELRIS QQGLDLIGNV EECRRDPYHC
 51 PADVLTVGIG STEANGKNID PKKRYSDKEI AQRMAYDLRL AEQCVARYGN
 101 GKNLPGARD AFSITFVNG CGKMOKSTLF KOANOGFTPO LCHOFRWIX
 151 AGGKTLGLV ARRAKALC LGETHD

IIAA_SEQUENCE 1.0
 ID R42374 standard; Protein: 178 AA.

AC R42374;
 DT 19-APR-1994 (first entry)
 DE Haemin-binding protein from ORF5 of plasmid pRAP501.
 KW Haemophilus somnus; immunogenic; haemolysin; LppB; LppC;
 KW thromboembolic meningoencephalitis; septicemia; arthritis;
 OS pneumonia; haemin-binding protein.
 OS Haemophilus somnus.

PN WO9321323-A.
 PD 28-OCT-1993.
 PF 05-APR-1993; CA0135.
 PR 09-APR-1992; US-865050.
 PR 04-JUN-1992; US-893424.
 PR 04-JUN-1992; US-893426.
 PR 29-MAR-1993; US-038287.
 PR 29-MAR-1993; US-038288.
 PR 29-MAR-1993; US-038719.
 PA (UYGA-) UNTV SASRATCHEMAN.
 PI Hatland RJ, Pfeiffer CG, Pontarollo RA, Potter AA;
 PI Rioux C, Theisen M;
 PI MPI: 93-351733/44.
 DR N-PSDB; Q51080.
 DR Haemophilus somnus immunogenic proteins used in vaccines -
 PT selected from haemin-binding protein, haemolysin, LppB and LppC,
 PT and corresp. DNA

PS Claim 1; Fig 2; 119pp; English.
 CC A genomic cosmid library of Haemophilus somnus HS25 DNA was screened
 CC for clones capable of binding bovine haemin and having haemolytic
 CC activity. Positive clones were subcloned various times, resulting
 CC in plasmid pRAP501, which binds haemin but is not haemolytic. The
 CC clone was sequenced and was found to contain several open reading
 CC frames, potentially encoding 8 proteins. The haemin binding protein
 CC (encoded by the hmb gene) was encoded by ORF1. The protein can be
 CC used in vaccines for preventing or treating H. somnus infections,
 CC which cause thromboembolic meningo-encephalitis, septicemia,
 CC arthritis and pneumonia in vertebrates.
 CC See also R42370-86.

CC Sequence 178 AA;
 SQ 16 A; 10 R; 11 N; 8 D; 0 B; 7 C; 11 Q; 9 E; 0 Z; 18 G; 4 H;
 SQ 10 I; 14 L; 15 K; 2 M; 7 F; 5 P; 6 S; 7 T; 3 W; 6 Y; 10 V;

R42374 Length: 178 April 28, 1998 12:48 Type: P Check: 4660 ..

1 MKFTFWGIG AICSVAILA LVKANOELR ISQOGLDLIG NVEGCRDPY
 51 HCPADVLTVG IGSTEANGKN IDPKRYSRDK EIAORWAYDL RLAEOCVNRY

101 GNGKLPQGA FDAFVSTFN VGCGKMOKST LFKOANOGFT POLCHOFERW
 151 IYAGKKLNG LVARRAKKA LCJGEYHD

IIAA_SEQUENCE 1.0
 ID R69760 standard; Protein: 112 AA.

AC R69760;
 DT 05-OCT-1995 (first entry)
 DE PMOS5973 (Met-15-125HuIL-3) having an M7 substitution.
 KW Hybrid; human; murine; interleukin 3; IL-3; deletion; substitution;
 KW biological activity; side effect profile; leukotriene; release;
 KW histamine; haematopoietic cell production; leucopenia.
 OS Homo sapiens.
 OS Mus musculus.

PN WO9500646-A.
 PD 05-JAN-1995.
 PF 17-JUN-1994; U06647.
 PR 21-JUN-1993; US-081539.
 PA (MONS) MONSANTO CO.
 PA (SEAR) SEARLE & CO G D.
 PI Bradford-Goldberg SR, Easton AM, Klein BK, McKearn JP;
 PI Ollins PO;
 PI MPI: 95-052083/07.
 DR N-PSDB; Q84028.

PT Human-murine interleukin 3 chimeric hybrid poly-peptide(s) -
 PT exhibit an improved side effects profile, e.g. reduction in
 PT stimulation of leukotriene or histamine release
 PS Claim 14; Page 126-127; 179pp; English.

CC The sequences given in R69752-64 represent hybrid proteins containing
 CC human and murine interleukin 3 (IL-3) segments. These proteins may
 CC contain deletions at the N- and/or C-terminus, and/or deletions/
 CC substitutions within the body of the protein. These proteins have
 CC at least the biological activity of native human IL-3 and have an
 CC improved side effect profile, eg. reduction in stimulation of
 CC leukotriene release or histamine release. They may be administered
 CC in a pharmaceutical composition to stimulate haematopoietic cell
 CC production in eg. leucopenia.
 CC Sequence 112 AA;
 SQ 8 A; 6 R; 13 N; 7 D; 0 B; 2 C; 6 Q; 8 E; 0 Z; 3 G; 3 H;
 SQ 8 I; 16 L; 5 K; 3 M; 4 F; 8 P; 4 S; 5 T; 1 W; 1 Y; 1 V;

R69760 Length: 112 April 28, 1998 12:48 Type: P Check: 1655 ..

1 MNCSSNIDEI ITHLKOPPLP LLDFFNNNGE DODILMENNUL RRPNDLEAFNR
 51 SGQEVNNSA IESILKNLPL CLPLATAAPT RHPHKKDD WNEFRKXLT
 101 YKLTENAOA QO

IIAA_SEQUENCE 1.0
 ID R78335 standard; Protein: 111 AA.

AC R78335;
 DT 27-JUN-1996 (first entry)
 DE Human IL-3 mutant for multi-lineage haematopoietic cell prodn.
 KW Interleukin-3; mutein; colony stimulating factor; CSF; chemotherapy;
 KW radiation; haematopoiesis; IL-3 antagonist; antigenic; antibody;
 OS synergy.
 OS Synthetic.

PN WO9520977-A1.
 PD 10-AUG-1995.
 PF 02-FEB-1995; U01184.
 PR 04-FEB-1994; US-193373.
 PA (SEAR) SEARLE & CO G D.
 PI Abrams MA, Bauer CS, Bradford-Goldberg SR, Caparon MH;
 PI Easton AM, Klein BK, McKearn JP, Ollins PO, Paik K;
 PI Thomas JM;
 PI MPI: 95-283609/37.

PT Multi-lineage hematopoietic cell production - by co-administration
 PT of human interleukin-3 variants and colony stimulating factors
 PT useful in radiation treatment and chemotherapy
 PS Claim 11; Page 249; 317pp; English.
 CC R78327-R78368 are human interleukin-3 (IL-3) mutant proteins
 CC (muteins). The IL-3 muteins are used in a method for increasing the

CC prodn. of multi-lineage haematopoietic cells in a mammal. The IL-3
 CC muteins are co-administered with a colony stimulating factor
 CC selected from GM-CSF, CSF-1, G-CSF, Meg-CSF, erythropoietin, IL-1,
 CC IL-2, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13,
 CC ILF, flt3/flt2, human growth hormone, B-cell growth factor,
 CC B-cell differentiation factor, eosinophil differentiation factor
 CC and stem cell factor. Co-administration enhances therapeutic value
 CC through synergistic effects. The mutant IL-3 molecules can be used to
 CC treat conditions in which haematopoietic cells have been depleted
 CC due to e.g. chemotherapy or radiation treatment. A further use of
 CC the IL-3 variants is as antagonists of IL-3 or antigenic fragments
 CC for the prodn. of antibodies.

Sequence 111 AA:

SQ 10 A: 7 R: 12 N: 5 D: 0 B: 2 C: 8 Q: 7 E: 0 Z: 3 G: 2 H:
 SQ 9 I: 15 L: 5 K: 2 M: 4 F: 8 P: 5 S: 4 T: 1 W: 1 Y: 1 V:

R78335 Length: 111 April 28, 1998 12:48 Type: P Check: 7559 ..

1 NCSNMIDEII THIKOPPLPI IDFNINLNGED QDIIMENLNR RPNLEAFNRA

51 VKSIQNASGI EAILRNLOPC LPSATAAPSR HPIIKAGDW QEFRRKLTFF

101 LKTIENNAOQ Q

!!AA_SEQUENCE 1.0

ID R56050 standard; peptide; 111 AA.

AC R56050;

DT 23-JUN-1995 (first entry)

DE Human interleukin-3 mutein (PMON13350).

KW Human interleukin-3; hIL-3; mutant; mutein.

OS Synthetic.

PN WO9412638-A.

PD 09-JUN-1994.

PF 22-NOV-1993: U11197.

PR 24-NOV-1992: US-981044.

PA (MONS) MONSANTO CO.

PI (SEAR) SEARLE & CO G D.

PI Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH;

PI Easton AM, Klein BK, McKearn JP, Olins PO, Palk K;

PI Thomas JW;

PI WPI: 94-200266/24.

DR N-PSDB; Q68098.

PT New human interleukin-3 mutant polypeptide(s) - useful for

PT treating various blood cell deficiencies, including leukaemia,

PT neutropenia and leukaemia

PS Claim 27: Page 90-91: 388pp: English.

CC The sequence is a mutein of hIL-3, comprising (15-125)hIL-3.

CC The region 15-125 comprises the substitutions: 73G, 76A, 79R, 82Q,

CC 87S, 93S, 98I, 101A, and 105Q. The muteins may have biological

CC activities similar to or better than hIL-3 and, in some cases, may

CC also have an improved side effect profile, e.g. better therapeutic

CC index. They may be used to treat e.g. leukaemia, neutropenia,

CC thrombocytopenia, (aplastic) anemia, Chediak-Higashi syndrome, SEF,

CC leukaemia, myelodysplastic syndrome, myelofibrosis; bone marrow

CC suppression or haematopoietic deficiency associated with treatment

CC with AZT or due to dialysis; or immunodeficiency resulting from viral

CC infection, exposure to recombinant DNA techniques. Recombinant DNA

CC sequences encoding the muteins are given in GENESEQ Accession Numbers

CC Q68098 - Q68189.

Sequence 111 AA:

SQ 10 A: 7 R: 12 N: 5 D: 0 B: 2 C: 8 Q: 7 E: 0 Z: 3 G: 2 H:
 SQ 9 I: 15 L: 5 K: 2 M: 4 F: 8 P: 5 S: 4 T: 1 W: 1 Y: 1 V;

R56050 Length: 111 April 28, 1998 12:48 Type: P Check: 7559 ..

1 NCSNMIDEII THIKOPPLPI IDFNINLNGED QDIIMENLNR RPNLEAFNRA

51 VKSIQNASGI EAILRNLOPC LPSATAAPSR HPIIKAGDW QEFRRKLTFF

101 LKTIENNAOQ Q

!!AA_SEQUENCE 1.0

ID P80381 standard; protein; 132 AA.

AC P80381;

DT 17-NOV-1990 (first entry)

DE Interleukin-3 variant.

KW Interleukin-3 variant; mutein;

KW colony stimulating factor activity.

OS Homo sapiens.

PN MO8806161-A.

PD 25-AUG-1988.

PF 18-FEB-1988: U00402.

PR 18-FEB-1987: US-016079.

PA (SCHB) Schering Biotech.

PI Otsuka T;

PI WPI: 88-249744/35.

PT Recombinant human interleukin-3 and muteins -

PT used for promoting growth and development of the hematopoietic

PT system and treating infection.

PS Disclosure; 4pp: English.

CC The human interleukin-3 variant or mutein exhibits colony

CC stimulating factor activity and comprises a glycosylated or

CC defined by AA 10-32, 47-60, 71-86 and 95-113. Inclusive, are

CC together not more than 3-fold substituted. The various

CC muteins stimulate the regeneration of blood cells and may be

CC useful in cancer therapy, treatment of certain blood

CC diseases and in the treatment of persistent infections.

CC See also P80382, P80383 and N80361.

SQ Sequence 132 AA:
 SQ 10 A: 6 R: 13 N: 6 D: 0 B: 2 C: 7 Q: 7 E: 0 Z: 2 G: 3 H:
 SQ 9 I: 20 L: 7 K: 3 M: 3 F: 12 P: 6 S: 11 T: 2 W: 1 Y: 2 V;

P80381 Length: 132 April 28, 1998 12:48 Type: P Check: 4868 ..

1 PMQTPTPLKT SWVNCNMID EITHLKOPP LPLIDPNLN GEDDILMEN

51 NLRPNLEAF NRAVKSQNA SAIESILNKL LPCUPLATAA PRRPHIRKD

101 GDWNEPRRL TFLKLENA QAQTTLSLA IF

!!AA_SEQUENCE 1.0

ID P80383 standard; protein; 133 AA.

AC P80383;

DT 18-NOV-1990 (first entry)

DE Interleukin-3 variant.

KW Interleukin-3 variant; mutein;

KW colony stimulating factor activity.

OS Homo sapiens.

PN MO8806161-A.

PD 25-AUG-1988.

PF 18-FEB-1988: U00402.

PR 18-FEB-1987: US-016079.

PA (SCHB) Schering Biotech.

PI Otsuka T;

PI WPI: 88-249744/35.

PT Recombinant human interleukin-3 and muteins -

PT used for promoting growth and development of the hematopoietic

PT system and treating infection.

PS Claim 13: Page 46-47: 54pp: English.

CC The human interleukin-3 variant or mutein exhibits colony

CC stimulating factor activity and comprises a glycosylated or

CC unglycosylated 10-fold substituted polypeptide where regions

CC defined by AA 10-32, 47-60, 71-86 and 95-113. Inclusive, are

CC together not more than 3-fold substituted. The various

CC muteins stimulate the regeneration of blood cells and may be

CC useful in cancer therapy, treatment of certain blood

CC diseases and in the treatment of persistent infections.

CC See also P80381, P80382 and N80361.

Sequence 133 AA:

SQ 11 A: 6 R: 13 N: 6 D: 0 B: 2 C: 7 Q: 7 E: 0 Z: 2 G: 3 H:
 SQ 9 I: 20 L: 7 K: 3 M: 3 F: 12 P: 6 S: 11 T: 2 W: 1 Y: 2 V;

P80383 Length: 133 April 28, 1998 12:48 Type: P Check: 6335 ..

```

1  ABMTQTPPK TSWWNCNNMI DEITTHKOP PLPLDPNNL NGEODITLME
51  NMLRRNLEA ENRAVKSJON ASAESILKN LPLCLPLATA APTRHPIHK
101  DGDWNEPRRK LTFYKLTEN AQAQOTTLST AIF

!!AA_SEQUENCE 1.0
ID  R56095 standard; peptide: 113 AA.
AC  R56095:
DE  26-JUN-1995 (first entry)
DE  Human interleukin-3 mutein (pMon13408).
KW  Human interleukin-3; hIL-3; mutant; mutein.
OS  Synthetic.
PN  WO9412638-A.
PD  09-JUN-1994.
PF  22-NOV-1993; U11197.
PR  24-NOV-1992; US-981044.
PA  (MONS ) MONSANTO CO.
PA  (SEAR ) SEARLE & CO G D.
PI  Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH;
PI  Easton AM, Klein BK, McKearn JP, Olin PO, Paik K;
PI  Thomas JW;
DR  N-PSDB; Q68121.
DR  WPI; 94-200266/24.
PT  New human interleukin-3 mutant polypeptide(s) - useful for
PT  treating various blood cell deficiencies, including leukaemia,
PT  neutropenia and leukaemia.
PS  Claim 32; Page 107-108; 388pp; English.
CC  The sequence is a mutein of hIL-3, comprising Met-Ala-(15-125)hIL-3.
CC  The muteins may have biological activities similar to or better than
CC  hIL-3 and, in some cases, may also have an improved side effect
CC  profile, e.g. better therapeutic index. They may be used to treat
CC  e.g. leukaemia, neutropenia, thrombocytopenia, (aplastic) anemia,
CC  Chediak-Higashi syndrome, SLE, leukaemia, myelodysplastic syndrome,
CC  myelofibrosis; bone marrow suppression or haematopoietic deficiency
CC  associated with treatment with AZT or due to dialysis; or
CC  immunodeficiency resulting from viral infection, exposure to
CC  radiation or cancer treatment. The muteins are made by recombinant
CC  DNA techniques. Recombinant DNA sequences encoding the muteins are
CC  given in GENESEQ Accession Numbers Q68083 - Q68189.
SO  Sequence 113 AA;
SQ  10 A; 7 R; 9 N; 5 D; 0 B; 2 C; 6 Q; 9 E; 0 Z; 2 G; 4 H;
SQ  11 I; 15 L; 4 K; 3 M; 3 F; 9 P; 5 S; 4 T; 1 W; 1 Y; 3 V;

R56095 Length: 113 April 28, 1998 12:48 Type: P Check: 6626 ..

1  MANCSIMIDE IHHUKRPPN PLDPNNLNS EDMDLIERN LRTPLILAFV
51  RAVKHLENAS GIEALIRNLQ PCUPSATAAP SRHPIIKAG DWQEFREKLT
101  FYIVTLEQAO EQQ

!!AA_SEQUENCE 1.0
ID  R56105 standard; peptide: 113 AA.
AC  R56105:
DE  26-JUN-1995 (first entry)
DE  Human interleukin-3 mutein (pMon13380).
KW  Human interleukin-3; hIL-3; mutant; mutein.
OS  Synthetic.
PN  WO9412638-A.
PD  09-JUN-1994.
PF  22-NOV-1993; U11197.
PR  24-NOV-1992; US-981044.
PA  (MONS ) MONSANTO CO.
PA  (SEAR ) SEARLE & CO G D.
PI  Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH;
PI  Easton AM, Klein BK, McKearn JP, Olin PO, Paik K;
PI  Thomas JW;
DR  N-PSDB; Q68131.
DR  WPI; 94-200266/24.
PT  New human interleukin-3 mutant polypeptide(s) - useful for
PT  treating various blood cell deficiencies, including leukaemia,

```

```

PT  neutropenia and leukaemia
PS  Claim 32; Page 112; 388pp; English.
CC  The sequence is a mutein of hIL-3, comprising Met-Ala-(15-125)hIL-3.
CC  The muteins may have biological activities similar to or better than
CC  hIL-3 and, in some cases, may also have an improved side effect
CC  profile, e.g. better therapeutic index. They may be used to treat
CC  e.g. leukaemia, neutropenia, thrombocytopenia, (aplastic) anemia,
CC  Chediak-Higashi syndrome, SLE, leukaemia, myelodysplastic syndrome,
CC  myelofibrosis; bone marrow suppression or haematopoietic deficiency
CC  associated with treatment with AZT or due to dialysis; or
CC  immunodeficiency resulting from viral infection, exposure to
CC  radiation or cancer treatment. The muteins are made by recombinant
CC  DNA techniques. Recombinant DNA sequences encoding the muteins are
CC  given in GENESEQ Accession Numbers Q68083 - Q68189.
SO  Sequence 113 AA;
SQ  12 A; 6 R; 9 N; 5 D; 0 B; 2 C; 7 Q; 9 E; 0 Z; 2 G; 3 H;
SQ  9 I; 16 L; 4 K; 4 M; 3 F; 8 P; 6 S; 2 T; 1 W; 1 Y; 4 V;

R56105 Length: 113 April 28, 1998 12:48 Type: P Check: 6426 ..

1  MANCSIMIDE AIHHUKVPPA PLDSNNLNS EDMDLIERN LRTPLILAFV
51  RAVKHLENAS GIEALIRNLQ PCUPSATAAP SRHPIIKAG DWQEFREKLT
101  FYIVTLEQAO EQQ

!!AA_SEQUENCE 1.0
ID  R56098 standard; peptide: 113 AA.
AC  R56098:
DE  26-JUN-1995 (first entry)
DE  Human interleukin-3 mutein (pMon13422).
KW  Human interleukin-3; hIL-3; mutant; mutein.
OS  Synthetic.
PN  WO9412638-A.
PD  09-JUN-1994.
PF  22-NOV-1993; U11197.
PR  24-NOV-1992; US-981044.
PA  (MONS ) MONSANTO CO.
PA  (SEAR ) SEARLE & CO G D.
PI  Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH;
PI  Easton AM, Klein BK, McKearn JP, Olin PO, Paik K;
PI  Thomas JW;
DR  N-PSDB; Q68124.
DR  WPI; 94-200266/24.
PT  New human interleukin-3 mutant polypeptide(s) - useful for
PT  treating various blood cell deficiencies, including leukaemia,
PT  neutropenia and leukaemia.
PS  Claim 32; Page 109; 388pp; English.
CC  The sequence is a mutein of hIL-3, comprising Met-Ala-(15-125)hIL-3.
CC  The muteins may have biological activities similar to or better than
CC  hIL-3 and, in some cases, may also have an improved side effect
CC  profile, e.g. better therapeutic index. They may be used to treat
CC  e.g. leukaemia, neutropenia, thrombocytopenia, (aplastic) anemia,
CC  Chediak-Higashi syndrome, SLE, leukaemia, myelodysplastic syndrome,
CC  myelofibrosis; bone marrow suppression or haematopoietic deficiency
CC  associated with treatment with AZT or due to dialysis; or
CC  immunodeficiency resulting from viral infection, exposure to
CC  radiation or cancer treatment. The muteins are made by recombinant
CC  DNA techniques. Recombinant DNA sequences encoding the muteins are
CC  given in GENESEQ Accession Numbers Q68083 - Q68189.
SO  Sequence 113 AA;
SQ  11 A; 7 R; 9 N; 5 D; 0 B; 2 C; 6 Q; 9 E; 0 Z; 2 G; 4 H;
SQ  11 I; 15 L; 4 K; 1 M; 3 F; 9 P; 5 S; 4 T; 1 W; 1 Y; 4 V;

R56098 Length: 113 April 28, 1998 12:48 Type: P Check: 6839 ..

1  MANCSIMIDE IHHUKRPPN PLDPNNLNS EDVDLIERN LRTPLILAFV
51  RAVKHLENAS GIEALIRNLQ PCUPSATAAP SRHPIIKAG DWQEFREKLT
101  FYIVTLEQAO EQQ

!!AA_SEQUENCE 1.0

```

ID R56099 standard; peptide: 113 AA.
AC R56099;
DT 26-JUN-1995 (first entry)
DE Human interleukin-3 mutein (pMon13423).
KW Human interleukin-3; hIL-3; mutant; mutein.
OS Synthetic.
PN WO9412638-A.
PD 09-JUN-1994.
PE 22-NOV-1993: U11197.
PR 24-NOV-1992: US-981044.
PA (MONS) MONSANTO CO.
PI (SEAR) SEARLE & CO G D.
PI Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH;
PI Easton AM, Klein BK, McKearn JP, Olin PO, Paik K;
PI Thomas JW;
DR WPI: 94-200266/24.
DR N-PSDB: Q68125.
PT New human interleukin-3 mutant polypeptide(s) - useful for
PT treating various blood cell deficiencies, including leukopenia,
PS neutropenia and leukaemia
PS Claim 32: Page 109-110: 388pp; English.
CC The sequence is a mutein of hIL-3, comprising Met-Ala-(15-125)hIL-3.
CC hIL-3 and, in some cases, may also have an improved side effect
CC profile, e.g. better therapeutic index. They may be used to treat
CC e.g. leukopenia, neutropenia, thrombocytopenia, (aplastic) anemia,
CC Chediak-Higashi syndrome, SLE, leukaemia, myelodysplastic syndrome,
CC myelofibrosis; bone marrow suppression or haematopoietic deficiency
CC associated with treatment with AZT or due to dialysis; or
CC immunodeficiency resulting from viral infection, exposure to
CC radiation or cancer treatment. The muteins are made by recombinant
CC DNA techniques. Recombinant DNA sequences encoding the muteins are
CC given in GENESPO Accession Numbers Q68083 - Q68189.
SQ Sequence 113 AA:
SQ 10 A; 7 R; 9 N; 5 D; 0 B; 2 C; 6 Q; 9 E; 0 Z; 2 G; 4 H;
SQ 12 I; 15 L; 4 K; 1 M; 3 F; 9 P; 5 S; 4 T; 1 W; 1 Y; 4 V;
R56099 Length: 113 April 28, 1998 12:48 Type: P Check: 6895 ..

1 MANCSIIDE IHHKRPNN PLDDNNINNS EDDILIERN LTRPNILAY
51 RAVKHEMNS GIEAILRNLQ PLPSAATAAP SRHPILIKAG DWQEFREKLT
101 FYLVLEQAO EQQ

11AA_SEQUENCE 1.0
ID W13009 standard; protein: 560 AA.
AC W13009;
DT 21-NOV-1997 (first entry)
DE Segment of desmosomal cadherin, desmoglein Dsg2.
KW Desmosomal cadherin; desmoglein; Dsg2; cell; surface; epithelial;
KW carcinoma; desmosome; antibody; epitope; diagnosis; detection;
KW micrometastasis; separation; enrichment; targeted delivery;
KW metastatic.
OS Homo sapiens.
PN DE19531033-A1.
PD 27-FEB-1997.
PE 23-AUG-1995: 031033.
PR 23-AUG-1995: DE-031033.
PA (PROG) PROGEN BIOTECHNIK GMBH.
PI Franke WW, Schaefer S;
DR WPI: 97-146518/14.
PT Antibody reactive with part of desmosomal cadherin - exposed on
PT surface of epithelial or carcinoma cells, not bound to desmosomes,
PT useful for diagnosis and treatment of carcinoma micrometastases
PS Claim 7: Page 5; 8pp; German.
CC The present sequence is a segment of the desmosomal cadherin (DC),
CC desmoglein Dsg2, which is exposed on the surface of epithelial or
CC carcinoma cells and not bound to desmosomes. An antibody (Ab)
CC directed against epitopes of the present sequence can be used to
CC diagnose, i.e. to detect carcinoma cells, especially
CC micrometastases, not bound to desmosomes, to separate, enrich or
CC detect living or fixed carcinoma cells by cell sorting methods and

CC as a therapeutic to deliver agents, e.g. other Ab or toxins, to
CC target cells. The Ab provides rapid and reliable detection of
CC metastatic carcinoma, and detects parts of DC that are not
CC accessible in desmosome bound cells, as in normal tissue or
CC carcinomas.
SQ Sequence 560 AA:
SQ 31 A; 19 R; 33 N; 39 D; 0 B; 5 C; 17 Q; 50 E; 0 Z; 32 G; 11 H;
SQ 43 I; 38 L; 38 K; 5 M; 22 F; 26 P; 39 S; 38 T; 4 W; 18 Y; 52 V;
W13009 Length: 560 April 28, 1998 12:48 Type: P Check: 3849 ..

1 AMITAPVALR EGEDLSKKNP IAKIHSDLAE ERGLKITRYK TSGKITEPFF
51 GIVFENKDG ELNVTSLIDR EETPFILTG YALDAGNNV EXPLELRITV
101 LDINDNEPVF TQDVVGSVE ELSAHTLVN KINATDADEP NTLNKSISR
151 IYSLERAYPP VFYLNKDTGE IYTVSVTLDR EEHSSTYLV EARDNGEYV
201 DKPVQAQVQ IRLDVNDNI PVVENKVLG MVENGVNVE VTRIVFDAD
251 EISDWMVLN PTFASGNEG YFHETDAQT NEGIVTLKE VDYEEKNID
301 FSVYANKAA FSDSRSRYK PTPIPIKVY KNYKEGIRK SSVISIVSE
351 SMDRSSKGI IGNFAPED TGLPAHARYV KLEDRDNWIS VDSVSEIKL
401 AKLPDEFNR VONGTYTYKI VAISEDYPRK TITGYLVIN EDINDCPTL
451 IEPVQITCD AEYVAVTAE LDGHPNSGPF SFSVIDKPG MAEKRIARQ
501 EESVVLQGS EKKLGRSEIQ FLISDNGFS CPBKQVLITL VCEVHGSGC
551 REAQHDSYVG

11AA_SEQUENCE 1.0
ID R56099 standard; peptide: 113 AA.
AC R56099;
DT 26-JUN-1995 (first entry)
DE Human interleukin-3 mutein (pMon13409).
KW Human interleukin-3; hIL-3; mutant; mutein.
OS Synthetic.
PN WO9412638-A.
PD 09-JUN-1994.
PE 22-NOV-1993: U11197.
PR 24-NOV-1992: US-981044.
PA (MONS) MONSANTO CO.
PI (SEAR) SEARLE & CO G D.
PI Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH;
PI Easton AM, Klein BK, McKearn JP, Olin PO, Paik K;
PI Thomas JW;
DR WPI: 94-200266/24.
DR N-PSDB: Q68122.
PT New human interleukin-3 mutant polypeptide(s) - useful for
PT treating various blood cell deficiencies, including leukopenia,
PT neutropenia and leukaemia
PS Claim 32: Page 108; 388pp; English.
CC The sequence is a mutein of hIL-3, comprising Met-Ala-(15-125)hIL-3.
CC hIL-3 and, in some cases, may also have an improved side effect
CC profile, e.g. better therapeutic index. They may be used to treat
CC e.g. leukopenia, neutropenia, thrombocytopenia, (aplastic) anemia,
CC Chediak-Higashi syndrome, SLE, leukaemia, myelodysplastic syndrome,
CC myelofibrosis; bone marrow suppression or haematopoietic deficiency
CC associated with treatment with AZT or due to dialysis; or
CC immunodeficiency resulting from viral infection, exposure to
CC radiation or cancer treatment. The muteins are made by recombinant
CC DNA techniques. Recombinant DNA sequences encoding the muteins are
CC given in GENESPO Accession Numbers Q68083 - Q68189.
SQ Sequence 113 AA:
SQ 10 A; 7 R; 9 N; 5 D; 0 B; 2 C; 6 Q; 9 E; 0 Z; 2 G; 4 H;
SQ 12 I; 15 L; 4 K; 3 M; 3 F; 9 P; 5 S; 4 T; 1 W; 1 Y; 3 V;

R56096 Length: 113 April 28, 1998 12:48 Type: P Check: 6737 ..

1 MANCSIMIDE IIHHLKRPPN PLLDPNNINS EDMOILMERN LRTPNLAFV

51 RAVKHEENAS GIEALIRNLQ PCLPSATAAP SRHPITIKAG DWQEFREKLT

101 FYLVTLQDAQ EQQ

11AA_SEQUENCE 1.0

ID R56090 standard; peptide: 113 AA.

AC R56090;

DE 26-JUN-1995 (first entry)

DE Human Interleukin-3 mutein (pMon13406).

KW Human Interleukin-3; hIL-3; mutant; mutein.

OS Synthetic.

PN MO9412638-A.

PD 09-JUN-1994.

PR 22-NOV-1993; U11197.

PR 24-NOV-1992; US-981044.

PA (MONS) MONSANTO CO.

PA (SEAR) SEARLE & CO G D.

PI Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH; Easton AM, Klein BK, McKearn JP, Olins PO, Paik K; Thomas JM; WPI: 94-200266/24.

DR N-PSDB; Q68116.

PT New human Interleukin-3 mutant polypeptide(s) - useful for treating various blood cell deficiencies, including leukopenia, neutropenia and leukaemia

PS Claim 32; Page 105-106; 388pp; English.

CC The sequence is a mutein of hIL-3, comprising Met-Ala-(15-125)hIL-3. The muteins may have biological activities similar to or better than hIL-3 and, in some cases, may also have an improved side effect profile, e.g. better therapeutic index. They may be used to treat e.g. leukopenia, neutropenia, thrombocytopenia, (aplastic) anemia, Chediak-Higashi syndrome, SLE, leukaemia, myelodysplastic syndrome, myelofibrosis; bone marrow suppression or haematopoietic deficiency associated with treatment with AZT or due to dialysis; or immunodeficiency resulting from viral infection, exposure to radiation or cancer treatment. The muteins are made by recombinant DNA techniques. Recombinant DNA sequences encoding the muteins are given in GENESQ Accession Numbers Q68083 - Q68189.

CC Sequence 113 AA:

SQ 11 A; 7 R; 9 N; 5 D; 0 B; 2 C; 6 Q; 9 E; 0 Z; 2 G; 4 H; 10 I; 15 L; 4 K; 3 M; 3 F; 9 P; 5 S; 4 T; 1 W; 1 Y; 3 V;

R56090 Length: 113 April 28, 1998 12:48 Type: P Check: 6690 ..

1 MANCSIAIDE IIHHLKRPPN PLLDPNNINS EDMOILMERN LRTPNLAFV

51 RAVKHEENAS GIEALIRNLQ PCLPSATAAP SRHPITIKAG DWQEFREKLT

101 FYLVTLQDAQ EQQ

11AA_SEQUENCE 1.0

ID R56091 standard; peptide: 113 AA.

AC R56091;

DE 26-JUN-1995 (first entry)

DE Human Interleukin-3 mutein (pMon13414).

KW Human Interleukin-3; hIL-3; mutant; mutein.

OS Synthetic.

PN MO9412638-A.

PD 09-JUN-1994.

PR 22-NOV-1993; U11197.

PR 24-NOV-1992; US-981044.

PA (MONS) MONSANTO CO.

PA (SEAR) SEARLE & CO G D.

PI Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH; Easton AM, Klein BK, McKearn JP, Olins PO, Paik K; Thomas JM; WPI: 94-200266/24.

DR N-PSDB; Q68118.

PT New human Interleukin-3 mutant polypeptide(s) - useful for

PT treating various blood cell deficiencies, including leukopenia, neutropenia and leukaemia

PS Claim 32; Page 106; 388pp; English.

CC The sequence is a mutein of hIL-3, comprising Met-Ala-(15-125)hIL-3. The muteins may have biological activities similar to or better than hIL-3 and, in some cases, may also have an improved side effect profile, e.g. better therapeutic index. They may be used to treat e.g. leukopenia, neutropenia, thrombocytopenia, (aplastic) anemia, Chediak-Higashi syndrome, SLE, leukaemia, myelodysplastic syndrome, myelofibrosis; bone marrow suppression or haematopoietic deficiency associated with treatment with AZT or due to dialysis; or immunodeficiency resulting from viral infection, exposure to radiation or cancer treatment. The muteins are made by recombinant DNA techniques. Recombinant DNA sequences encoding the muteins are given in GENESQ Accession Numbers Q68083 - Q68189.

CC Sequence 113 AA:

SQ 10 A; 7 R; 9 N; 5 D; 0 B; 2 C; 6 Q; 9 E; 0 Z; 2 G; 4 H; 11 I; 15 L; 4 K; 3 M; 3 F; 9 P; 5 S; 4 T; 1 W; 1 Y; 3 V;

R56091 Length: 113 April 28, 1998 12:48 Type: P Check: 6746 ..

1 MANCSIIIDE IIHHLKRPPN PLLDPNNINS EDMOILMERN LRTPNLAFV

51 RAVKHEENAS GIEALIRNLQ PCLPSATAAP SRHPITIKAG DWQEFREKLT

101 FYLVTLQDAQ EQQ

11AA_SEQUENCE 1.0

ID R78352;

AC R78352;

DE 27-JUN-1996 (first entry)

DE Human IL-3 mutant for multi-lineage haematopoietic cell prodn.

KW Interleukin-3; mutein; colony stimulating factor; CSF; chemotherapy; radiation; haematopoiesis; IL-3 antagonist; antigenic; antibody;

KW synergy.

OS Synthetic.

PN WO9520977-A1.

PD 10-AUG-1995.

PR 04-FEB-1995; U01184.

PR 04-FEB-1994; US-193373.

PA (SEAR) SEARLE & CO G D.

PI Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH; Easton AM, Klein BK, McKearn JP, Olins PO, Paik K; Thomas JM; WPI: 95-283609/37.

DR WPI: 95-283609/37.

PT Multi-lineage hematopoietic cell production - by co-administration of human Interleukin-3 variants and colony stimulating factors

PT useful in radiation treatment and chemotherapy

PS Claim 11; Page 253-254; 317pp; English.

CC R78327-R78368 are human interleukin-3 (IL-3) mutant proteins (muteins). The IL-3 muteins are used in a method for increasing the prodn. of multi-lineage haematopoietic cells in a mammal. The IL-3 muteins are co-administered with a colony stimulating factor selected from GM-CSF, CSF-1, G-CSF, Meg-CSF, erythropoietin, IL-1, IL-2, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, IL-16, IL-17, IL-18, IL-19, IL-20, IL-21, IL-22, IL-23, IL-24, IL-25, IL-26, IL-27, IL-28, IL-29, IL-30, IL-31, IL-32, IL-33, IL-34, IL-35, IL-36, IL-37, IL-38, IL-39, IL-40, IL-41, IL-42, IL-43, IL-44, IL-45, IL-46, IL-47, IL-48, IL-49, IL-50, IL-51, IL-52, IL-53, IL-54, IL-55, IL-56, IL-57, IL-58, IL-59, IL-60, IL-61, IL-62, IL-63, IL-64, IL-65, IL-66, IL-67, IL-68, IL-69, IL-70, IL-71, IL-72, IL-73, IL-74, IL-75, IL-76, IL-77, IL-78, IL-79, IL-80, IL-81, IL-82, IL-83, IL-84, IL-85, IL-86, IL-87, IL-88, IL-89, IL-90, IL-91, IL-92, IL-93, IL-94, IL-95, IL-96, IL-97, IL-98, IL-99, IL-100, IL-101, IL-102, IL-103, IL-104, IL-105, IL-106, IL-107, IL-108, IL-109, IL-110, IL-111, IL-112, IL-113, IL-114, IL-115, IL-116, IL-117, IL-118, IL-119, IL-120, IL-121, IL-122, IL-123, IL-124, IL-125, IL-126, IL-127, IL-128, IL-129, IL-130, IL-131, IL-132, IL-133, IL-134, IL-135, IL-136, IL-137, IL-138, IL-139, IL-140, IL-141, IL-142, IL-143, IL-144, IL-145, IL-146, IL-147, IL-148, IL-149, IL-150, IL-151, IL-152, IL-153, IL-154, IL-155, IL-156, IL-157, IL-158, IL-159, IL-160, IL-161, IL-162, IL-163, IL-164, IL-165, IL-166, IL-167, IL-168, IL-169, IL-170, IL-171, IL-172, IL-173, IL-174, IL-175, IL-176, IL-177, IL-178, IL-179, IL-180, IL-181, IL-182, IL-183, IL-184, IL-185, IL-186, IL-187, IL-188, IL-189, IL-190, IL-191, IL-192, IL-193, IL-194, IL-195, IL-196, IL-197, IL-198, IL-199, IL-200, IL-201, IL-202, IL-203, IL-204, IL-205, IL-206, IL-207, IL-208, IL-209, IL-210, IL-211, IL-212, IL-213, IL-214, IL-215, IL-216, IL-217, IL-218, IL-219, IL-220, IL-221, IL-222, IL-223, IL-224, IL-225, IL-226, IL-227, IL-228, IL-229, IL-230, IL-231, IL-232, IL-233, IL-234, IL-235, IL-236, IL-237, IL-238, IL-239, IL-240, IL-241, IL-242, IL-243, IL-244, IL-245, IL-246, IL-247, IL-248, IL-249, IL-250, IL-251, IL-252, IL-253, IL-254, IL-255, IL-256, IL-257, IL-258, IL-259, IL-260, IL-261, IL-262, IL-263, IL-264, IL-265, IL-266, IL-267, IL-268, IL-269, IL-270, IL-271, IL-272, IL-273, IL-274, IL-275, IL-276, IL-277, IL-278, IL-279, IL-280, IL-281, IL-282, IL-283, IL-284, IL-285, IL-286, IL-287, IL-288, IL-289, IL-290, IL-291, IL-292, IL-293, IL-294, IL-295, IL-296, IL-297, IL-298, IL-299, IL-300, IL-301, IL-302, IL-303, IL-304, IL-305, IL-306, IL-307, IL-308, IL-309, IL-310, IL-311, IL-312, IL-313, IL-314, IL-315, IL-316, IL-317, IL-318, IL-319, IL-320, IL-321, IL-322, IL-323, IL-324, IL-325, IL-326, IL-327, IL-328, IL-329, IL-330, IL-331, IL-332, IL-333, IL-334, IL-335, IL-336, IL-337, IL-338, IL-339, IL-340, IL-341, IL-342, IL-343, IL-344, IL-345, IL-346, IL-347, IL-348, IL-349, IL-350, IL-351, IL-352, IL-353, IL-354, IL-355, IL-356, IL-357, IL-358, IL-359, IL-360, IL-361, IL-362, IL-363, IL-364, IL-365, IL-366, IL-367, IL-368, IL-369, IL-370, IL-371, IL-372, IL-373, IL-374, IL-375, IL-376, IL-377, IL-378, IL-379, IL-380, IL-381, IL-382, IL-383, IL-384, IL-385, IL-386, IL-387, IL-388, IL-389, IL-390, IL-391, IL-392, IL-393, IL-394, IL-395, IL-396, IL-397, IL-398, IL-399, IL-400, IL-401, IL-402, IL-403, IL-404, IL-405, IL-406, IL-407, IL-408, IL-409, IL-410, IL-411, IL-412, IL-413, IL-414, IL-415, IL-416, IL-417, IL-418, IL-419, IL-420, IL-421, IL-422, IL-423, IL-424, IL-425, IL-426, IL-427, IL-428, IL-429, IL-430, IL-431, IL-432, IL-433, IL-434, IL-435, IL-436, IL-437, IL-438, IL-439, IL-440, IL-441, IL-442, IL-443, IL-444, IL-445, IL-446, IL-447, IL-448, IL-449, IL-450, IL-451, IL-452, IL-453, IL-454, IL-455, IL-456, IL-457, IL-458, IL-459, IL-460, IL-461, IL-462, IL-463, IL-464, IL-465, IL-466, IL-467, IL-468, IL-469, IL-470, IL-471, IL-472, IL-473, IL-474, IL-475, IL-476, IL-477, IL-478, IL-479, IL-480, IL-481, IL-482, IL-483, IL-484, IL-485, IL-486, IL-487, IL-488, IL-489, IL-490, IL-491, IL-492, IL-493, IL-494, IL-495, IL-496, IL-497, IL-498, IL-499, IL-500, IL-501, IL-502, IL-503, IL-504, IL-505, IL-506, IL-507, IL-508, IL-509, IL-510, IL-511, IL-512, IL-513, IL-514, IL-515, IL-516, IL-517, IL-518, IL-519, IL-520, IL-521, IL-522, IL-523, IL-524, IL-525, IL-526, IL-527, IL-528, IL-529, IL-530, IL-531, IL-532, IL-533, IL-534, IL-535, IL-536, IL-537, IL-538, IL-539, IL-540, IL-541, IL-542, IL-543, IL-544, IL-545, IL-546, IL-547, IL-548, IL-549, IL-550, IL-551, IL-552, IL-553, IL-554, IL-555, IL-556, IL-557, IL-558, IL-559, IL-560, IL-561, IL-562, IL-563, IL-564, IL-565, IL-566, IL-567, IL-568, IL-569, IL-570, IL-571, IL-572, IL-573, IL-574, IL-575, IL-576, IL-577, IL-578, IL-579, IL-580, IL-581, IL-582, IL-583, IL-584, IL-585, IL-586, IL-587, IL-588, IL-589, IL-590, IL-591, IL-592, IL-593, IL-594, IL-595, IL-596, IL-597, IL-598, IL-599, IL-600, IL-601, IL-602, IL-603, IL-604, IL-605, IL-606, IL-607, IL-608, IL-609, IL-610, IL-611, IL-612, IL-613, IL-614, IL-615, IL-616, IL-617, IL-618, IL-619, IL-620, IL-621, IL-622, IL-623, IL-624, IL-625, IL-626, IL-627, IL-628, IL-629, IL-630, IL-631, IL-632, IL-633, IL-634, IL-635, IL-636, IL-637, IL-638, IL-639, IL-640, IL-641, IL-642, IL-643, IL-644, IL-645, IL-646, IL-647, IL-648, IL-649, IL-650, IL-651, IL-652, IL-653, IL-654, IL-655, IL-656, IL-657, IL-658, IL-659, IL-660, IL-661, IL-662, IL-663, IL-664, IL-665, IL-666, IL-667, IL-668, IL-669, IL-670, IL-671, IL-672, IL-673, IL-674, IL-675, IL-676, IL-677, IL-678, IL-679, IL-680, IL-681, IL-682, IL-683, IL-684, IL-685, IL-686, IL-687, IL-688, IL-689, IL-690, IL-691, IL-692, IL-693, IL-694, IL-695, IL-696, IL-697, IL-698, IL-699, IL-700, IL-701, IL-702, IL-703, IL-704, IL-705, IL-706, IL-707, IL-708, IL-709, IL-710, IL-711, IL-712, IL-713, IL-714, IL-715, IL-716, IL-717, IL-718, IL-719, IL-720, IL-721, IL-722, IL-723, IL-724, IL-725, IL-726, IL-727, IL-728, IL-729, IL-730, IL-731, IL-732, IL-733, IL-734, IL-735, IL-736, IL-737, IL-738, IL-739, IL-740, IL-741, IL-742, IL-743, IL-744, IL-745, IL-746, IL-747, IL-748, IL-749, IL-750, IL-751, IL-752, IL-753, IL-754, IL-755, IL-756, IL-757, IL-758, IL-759, IL-760, IL-761, IL-762, IL-763, IL-764, IL-765, IL-766, IL-767, IL-768, IL-769, IL-770, IL-771, IL-772, IL-773, IL-774, IL-775, IL-776, IL-777, IL-778, IL-779, IL-780, IL-781, IL-782, IL-783, IL-784, IL-785, IL-786, IL-787, IL-788, IL-789, IL-790, IL-791, IL-792, IL-793, IL-794, IL-795, IL-796, IL-797, IL-798, IL-799, IL-800, IL-801, IL-802, IL-803, IL-804, IL-805, IL-806, IL-807, IL-808, IL-809, IL-810, IL-811, IL-812, IL-813, IL-814, IL-815, IL-816, IL-817, IL-818, IL-819, IL-820, IL-821, IL-822, IL-823, IL-824, IL-825, IL-826, IL-827, IL-828, IL-829, IL-830, IL-831, IL-832, IL-833, IL-834, IL-835, IL-836, IL-837, IL-838, IL-839, IL-840, IL-841, IL-842, IL-843, IL-844, IL-845, IL-846, IL-847, IL-848, IL-849, IL-850, IL-851, IL-852, IL-853, IL-854, IL-855, IL-856, IL-857, IL-858, IL-859, IL-860, IL-861, IL-862, IL-863, IL-864, IL-865, IL-866, IL-867, IL-868, IL-869, IL-870, IL-871, IL-872, IL-873, IL-874, IL-875, IL-876, IL-877, IL-878, IL-879, IL-880, IL-881, IL-882, IL-883, IL-884, IL-885, IL-886, IL-887, IL-888, IL-889, IL-890, IL-891, IL-892, IL-893, IL-894, IL-895, IL-896, IL-897, IL-898, IL-899, IL-900, IL-901, IL-902, IL-903, IL-904, IL-905, IL-906, IL-907, IL-908, IL-909, IL-910, IL-911, IL-912, IL-913, IL-914, IL-915, IL-916, IL-917, IL-918, IL-919, IL-920, IL-921, IL-922, IL-923, IL-924, IL-925, IL-926, IL-927, IL-928, IL-929, IL-930, IL-931, IL-932, IL-933, IL-934, IL-935, IL-936, IL-937, IL-938, IL-939, IL-940, IL-941, IL-942, IL-943, IL-944, IL-945, IL-946, IL-947, IL-948, IL-949, IL-950, IL-951, IL-952, IL-953, IL-954, IL-955, IL-956, IL-957, IL-958, IL-959, IL-960, IL-961, IL-962, IL-963, IL-964, IL-965, IL-966, IL-967, IL-968, IL-969, IL-970, IL-971, IL-972, IL-973, IL-974, IL-975, IL-976, IL-977, IL-978, IL-979, IL-980, IL-981, IL-982, IL-983, IL-984, IL-985, IL-986, IL-987, IL-988, IL-989, IL-990, IL-991, IL-992, IL-993, IL-994, IL-995, IL-996, IL-997, IL-998, IL-999, IL-1000, IL-1001, IL-1002, IL-1003, IL-1004, IL-1005, IL-1006, IL-1007, IL-1008, IL-1009, IL-1010, IL-1011, IL-1012, IL-1013, IL-1014, IL-1015, IL-1016, IL-1017, IL-1018, IL-1019, IL-1020, IL-1021, IL-1022, IL-1023, IL-1024, IL-1025, IL-1026, IL-1027, IL-1028, IL-1029, IL-1030, IL-1031, IL-1032, IL-1033, IL-1034, IL-1035, IL-1036, IL-1037, IL-1038, IL-1039, IL-1040, IL-1041, IL-1042, IL-1043, IL-1044, IL-1045, IL-1046, IL-1047, IL-1048, IL-1049, IL-1050, IL-1051, IL-1052, IL-1053, IL-1054, IL-1055, IL-1056, IL-1057, IL-1058, IL-1059, IL-1060, IL-1061, IL-1062, IL-1063, IL-1064, IL-1065, IL-1066, IL-1067, IL-1068, IL-1069, IL-1070, IL-1071, IL-1072, IL-1073, IL-1074, IL-1075, IL-1076, IL-1077, IL-1078, IL-1079, IL-1080, IL-1081, IL-1082, IL-1083, IL-1084, IL-1085, IL-1086, IL-1087, IL-1088, IL-1089, IL-1090, IL-1091, IL-1092, IL-1093, IL-1094, IL-1095, IL-1096, IL-1097, IL-1098, IL-1099, IL-1100, IL-1101, IL-1102, IL-1103, IL-1104, IL-1105, IL-1106, IL-1107, IL-1108, IL-1109, IL-1110, IL-1111, IL-1112, IL-1113, IL-1114, IL-1115, IL-1116, IL-1117, IL-1118, IL-1119, IL-1120, IL-1121, IL-1122, IL-1123, IL-1124, IL-1125, IL-1126, IL-1127, IL-1128, IL-1129, IL-1130, IL-1131, IL-1132, IL-1133, IL-1134, IL-1135, IL-1136, IL-1137, IL-1138, IL-1139, IL-1140, IL-1141, IL-1142, IL-1143, IL-1144, IL-1145, IL-1146, IL-1147, IL-1148, IL-1149, IL-1150, IL-1151, IL-1152, IL-1153, IL-1154, IL-1155, IL-1156, IL-1157, IL-1158, IL-1159, IL-1160, IL-1161, IL-1162, IL-1163, IL-1164, IL-1165, IL-1166, IL-1167, IL-1168, IL-1169, IL-1170, IL-1171, IL-1172, IL-1173, IL-1174, IL-1175, IL-1176, IL-1177, IL-1178, IL-1179, IL-1180, IL-1181, IL-1182, IL-1183, IL-1184, IL-1185, IL-1186, IL-1187, IL-1188, IL-1189, IL-1190, IL-1191, IL-1192, IL-1193, IL-1194, IL-1195, IL-1196, IL-1197, IL-1198, IL-1199, IL-1200, IL-1201, IL-1202, IL-1203, IL-1204, IL-1205, IL-1206, IL-1207, IL-1208, IL-1209, IL-1210, IL-1211, IL-1212, IL-1213, IL-1214, IL-1215, IL-1216, IL-1217, IL-1218, IL-1219, IL-1220, IL-1221, IL-1222, IL-1223, IL-1224, IL-1225, IL-1226, IL-1227, IL-1228, IL-1229, IL-1230, IL-1231, IL-1232, IL-1233, IL-1234, IL-1235, IL-1236, IL-1237, IL-1238, IL-1239, IL-1240, IL-1241, IL-1242, IL-1243, IL-1244, IL-1245, IL-1246, IL-1247, IL-1248, IL-1249, IL-1250, IL-1251, IL-1252, IL-1253, IL-1254, IL-1255, IL-1256, IL-1257, IL-1258, IL-1259, IL-1260, IL-1261, IL-1262, IL-1263, IL-1264, IL-1265, IL-1266, IL-1267, IL-1268, IL-1269, IL-1270, IL-1271, IL-1272, IL-1273, IL-1274, IL-1275, IL-1276, IL-1277, IL-1278, IL-1279, IL-1280, IL-1281, IL-1282, IL-1283, IL-1284, IL-1285, IL-1286, IL-1287, IL-1288, IL-1289, IL-1290, IL-1291, IL-1292, IL-1293, IL-1294, IL-1295, IL-1296, IL-1297, IL-1298, IL-1299, IL-1300, IL-1301, IL-1302, IL-1303, IL-1304, IL-1305, IL-1306, IL-1307, IL-1308, IL-1309, IL-1310, IL-1311, IL-1312, IL-1313, IL-1314, IL-1315, IL-1316, IL-1317, IL-1318, IL-1319, IL-1320, IL-1321, IL-1322, IL-1323, IL-1324, IL-1325, IL-1326, IL-1327, IL-1328, IL-1329, IL-1330, IL-1331, IL-1332, IL-1333, IL-1334, IL-1335, IL-1336, IL-1337, IL-1338, IL-1339, IL-1340, IL-1341, IL-1342, IL-1343, IL-1344, IL-1345, IL-1346, IL-1347, IL-1348, IL-1349, IL-1350, IL-1351, IL-1352, IL-1353, IL-1354, IL-1355, IL-1356, IL-1357, IL-1358, IL-1359, IL-1360, IL-1361, IL-1362, IL-1363, IL-1364, IL-1365, IL-1366, IL-1367, IL-1368, IL-1369, IL-1370, IL-1371, IL-1372, IL-1373, IL-1374, IL-1375, IL-1376, IL-1377, IL-1378, IL-1379, IL-1380, IL-1381, IL-1382, IL-1383, IL-1384, IL-1385, IL-1386, IL-1387, IL-1388, IL-1389, IL-1390, IL-1391, IL-1392, IL-1393, IL-1394, IL-1395, IL-1396, IL-1397, IL-1398, IL-1399, IL-1400, IL-1401, IL-1402, IL-1403, IL-1404, IL-1405, IL-1406, IL-1407, IL-1408, IL-1409, IL-1410, IL-1411, IL-1412, IL-1413, IL-1414, IL-1415, IL-1416, IL-1417, IL-1418, IL-1419, IL-1420, IL-1421, IL-1422, IL-1423, IL-1424, IL-1425, IL-1426, IL-1427, IL-1428, IL-1429, IL-1430, IL-1431, IL-1432, IL-1433, IL-1434, IL-1435, IL-1436, IL-1437, IL-1438, IL-1439, IL-1440, IL-1441, IL-1442, IL-1443, IL-1444, IL-1445, IL-1446, IL-1447, IL-1448, IL-1449, IL-1450, IL-1451, IL-1452, IL-1453, IL-1454, IL-1455, IL-1456, IL-1457, IL-1458, IL-1459, IL-1460, IL-1461, IL-1462, IL-1463, IL-1464, IL-1465, IL-1466, IL-1467, IL-1468, IL-1469, IL-1470, IL-1471, IL-1472, IL-1473, IL-1474, IL-1475, IL-1476, IL-1477, IL-1478, IL-1479, IL-1480, IL-1481, IL-1482, IL-1483, IL-1484, IL-1485, IL-1486, IL-1487, IL-1488, IL-1489, IL-1490, IL-1491, IL-1492, IL-1493, IL-1494, IL-1495, IL-1496, IL-1497, IL-1498, IL-1499, IL-1500, IL-1501, IL-1502, IL-1503, IL-1504, IL-1505, IL-1506, IL-1507, IL-1508, IL-1509, IL-1510, IL-1511, IL-1512, IL-1513, IL-1514, IL-1515, IL-1516, IL-1517, IL-1518, IL-1519, IL-1520, IL-1521, IL-1522, IL-1523, IL-1524, IL-1525, IL-1526, IL-1527, IL-1528, IL-1529, IL-1530, IL-1531, IL-1532, IL-1533, IL-1534, IL-1535, IL-1536, IL-1537, IL-1538, IL-1539, IL-1540, IL-1541, IL-1542, IL-1543, IL-1544, IL-1545, IL-1546, IL-1547, IL-1548, IL-1549, IL-1550, IL-1551, IL-1552, IL-1553, IL-1554, IL-1555, IL-1556, IL-1557, IL-1558, IL-1559, IL-1560, IL-1561, IL-1562, IL-1563, IL-1564, IL-1565, IL-1566, IL-1567, IL-1568, IL-

```

101 FYLVLEQAO EQQ
!!AA_SEQUENCE 1.0
ID R56137 standard; peptide; 113 AA.
AC R56137;
DT 26-JUN-1995 (first entry)
DE Human interleukin-3 mutein (pMon13440).
KW Human interleukin-3; hIL-3; mutant; mutein.
OS Synthetic.
PN MO9412638-A.
PD 09-JUN-1994.
PE 22-NOV-1993; U11197.
PR 24-NOV-1992; US-981044.
PA (MONS ) MONSANTO CO.
PA (SEAR ) SEARLE & CO G D.
PI Abrams MA, Bauer SC, Bratford-Goldberg SR, Caparon MH;
PI Easton AM, Klein BK, McKearn JP, Olins PO, Paik K;
PI Thomas JW;
DR N-PSDB; Q68170.
DR WPI; 94-200266/24.
PT New human interleukin-3 mutant polypeptide(s) - useful for
PT treating various blood cell deficiencies, including leukopenia,
PT neutropenia and leukaemia
PS Claim 32; Page 130-131; 388pp; English.
CC The sequence is a mutein of hIL-3, comprising Met-Ala-(15-125)hIL-3.
CC The muteins may have biological activities similar to or better than
CC hIL-3 and, in some cases, may also have an improved side effect
CC profile, e.g. better therapeutic index. They may be used to treat
CC e.g. leukopenia, neutropenia, thrombocytopenia, (aplastic) anemia,
CC Chediak-Higashi syndrome, SLE, leukaemia, myelodysplastic syndrome,
CC myelofibrosis; bone marrow suppression or haematopoietic deficiency
CC associated with treatment with AZT or due to dialysis; or
CC immunodeficiency resulting from viral infection, exposure to
CC radiation or cancer treatment. The muteins are made by recombinant
CC DNA techniques. Recombinant DNA sequences encoding the muteins are
CC given in GENESQ Accession Numbers Q68083 - Q68189.
SQ Sequence 113 AA;
SQ 12 A; 6 R; 8 N; 4 D; 0 B; 2 C; 6 O; 9 E; 0 Z; 2 G; 4 H;
SQ 10 I; 14 L; 4 K; 3 M; 3 F; 8 P; 8 S; 4 T; 1 W; 1 Y; 4 V;
SQ 10 I; 14 L; 4 K; 3 M; 3 F; 8 P; 8 S; 4 T; 1 W; 1 Y; 4 V;
R56137 Length: 113 April 28, 1998 12:48 Type: P Check: 7231 ..

1 MANCSIMIDE AIHRLKIPPN PSIDSANLNS EDVSIEMERN LRTPNLAFV
51 RAVKLENAS GIEALIRNLQ PCLPSTAP SRHPIITKAG DMQERREKLT
101 FYLVLEQAO EQQ

!!AA_SEQUENCE 1.0
ID R78368 standard; protein; 113 AA.
AC R78368;
DT 27-JUN-1996 (first entry)
DE Human IL-3 mutant for multi-lineage haematopoietic cell prodn.
KW Interleukin-3; mutein; colony stimulating factor; CSF; chemotherapy;
KW radiation; haematopoiesis; IL-3 antagonist; antigenic; antibody;
OS Synthetic.
PN MO9520977-A1.
PD 10-AUG-1995.
PE 02-FEB-1995; U01184.
PR 04-FEB-1994; US-193373.
PA (SEAR ) SEARLE & CO G D.
PI Abrams MA, Bauer SC, Bratford-Goldberg SR, Caparon MH;
PI Easton AM, Klein BK, McKearn JP, Olins PO, Paik K;
PI Thomas JW;
DR WPI; 95-283609/37.
PT Multi-lineage hematopoietic cell production - by co-administration
PT of human interleukin-3 variants and colony stimulating factors
PT useful in radiation treatment and chemotherapy
PS Claim 12; Page 258; 317pp; English.
CC R78327-R78368 are human interleukin-3 (IL-3) mutant proteins
CC (muteins). The IL-3 muteins are used in a method for increasing the
CC prodn. of multi-lineage haematopoietic cells in a mammal. The IL-3

```

```

CC muteins are co-administered with a colony stimulating factor
CC selected from GM-CSF, CSF-1, G-CSF, Meg-CSF, erythropoietin, IL-1,
CC IL-2, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13,
CC LIF, flt3/flt2, human growth hormone, B-cell growth factor,
CC B-cell differentiation factor, eosinophil differentiation factor
CC and stem cell factor. Co-administration enhances therapeutic value
CC through synergistic effects. The mutant IL-3 molecules can be used to
CC treat conditions in which haematopoietic cells have been depleted
CC due to e.g. chemotherapy or radiation treatment. A further use of
CC the IL-3 variants is as antagonists of IL-3 or antigenic fragments
CC for the prodn. of antibodies.
SQ Sequence 113 AA;
SQ 11 A; 6 R; 9 N; 4 D; 0 B; 2 C; 6 O; 9 E; 0 Z; 2 G; 4 H;
SQ 9 I; 15 L; 4 K; 3 M; 3 F; 8 P; 8 S; 4 T; 1 W; 1 Y; 4 V;
SQ 9 I; 15 L; 4 K; 3 M; 3 F; 8 P; 8 S; 4 T; 1 W; 1 Y; 4 V;
R78368 Length: 113 April 28, 1998 12:48 Type: P Check: 7382 ..

1 MANCSIMIDE LIHRLKIPPN PSIDSANLNS EDVSIEMERN LRTPNLAFV
51 RAVKLENAS GIEALIRNLQ PCLPSTAP SRHPIITKAG DMQERREKLT
101 FYLVLEQAO EQQ

!!AA_SEQUENCE 1.0
ID R56120 standard; peptide; 113 AA.
AC R56120;
DT 26-JUN-1995 (first entry)
DE Human interleukin-3 mutein (pMon13383).
KW Human interleukin-3; hIL-3; mutant; mutein.
OS Synthetic.
PN MO9412638-A.
PD 09-JUN-1994.
PE 22-NOV-1993; U11197.
PR 24-NOV-1992; US-981044.
PA (MONS ) MONSANTO CO.
PA (SEAR ) SEARLE & CO G D.
PI Abrams MA, Bauer SC, Bratford-Goldberg SR, Caparon MH;
PI Easton AM, Klein BK, McKearn JP, Olins PO, Paik K;
PI Thomas JW;
DR N-PSDB; Q68146.
DR WPI; 94-200266/24.
PT New human interleukin-3 mutant polypeptide(s) - useful for
PT treating various blood cell deficiencies, including leukopenia,
PT neutropenia and leukaemia
PS Disclosure; Page 119; 388pp; English.
CC The sequence is a mutein of hIL-3, comprising Met-Ala-(15-125)hIL-3.
CC The muteins may have biological activities similar to or better than
CC hIL-3 and, in some cases, may also have an improved side effect
CC profile, e.g. better therapeutic index. They may be used to treat
CC e.g. leukopenia, neutropenia, thrombocytopenia, (aplastic) anemia,
CC Chediak-Higashi syndrome, SLE, leukaemia, myelodysplastic syndrome,
CC myelofibrosis; bone marrow suppression or haematopoietic deficiency
CC associated with treatment with AZT or due to dialysis; or
CC immunodeficiency resulting from viral infection, exposure to
CC radiation or cancer treatment. The muteins are made by recombinant
CC DNA techniques. Recombinant DNA sequences encoding the muteins are
CC given in GENESQ Accession Numbers Q68083 - Q68189.
SQ Sequence 113 AA;
SQ 11 A; 7 R; 9 N; 6 D; 0 B; 2 C; 7 O; 9 E; 0 Z; 2 G; 3 H;
SQ 9 I; 15 L; 4 K; 3 M; 3 F; 9 P; 6 S; 2 T; 1 W; 1 Y; 4 V;
SQ 9 I; 15 L; 4 K; 3 M; 3 F; 9 P; 6 S; 2 T; 1 W; 1 Y; 4 V;
R56120 Length: 113 April 28, 1998 12:48 Type: P Check: 7137 ..

1 MANCSIMIDE AIHRLKIPPN PLIDPNLND EDVSIEMERN LRLPNLESFV
51 RAVKLENAS GIEALIRNLQ PCLPSTAP SRHPIITKAG DMQERREKLT
101 FYLVLEQAO EQQ

!!AA_SEQUENCE 1.0
ID R56067 standard; peptide; 113 AA.
AC R56067;
DT 22-JUN-1995 (first entry)

```

DE Human interleukin-3 mutein (pMon13288).
 KW Human interleukin-3; hIL-3; mutant; mutein.
 OS Synthetic.
 PN W09412638-A.
 PD 09-JUN-1994.
 PF 22-NOV-1993: U11197.
 PR 24-NOV-1992: US-981044.
 PA (MONS) MONSANTO CO.
 PI Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH, Easton AM, Klein BK, McKeam JP, Olins PO, Paik K; Thomas JW; WPI: 94-200266/24.
 DR N-PSDB; Q68086.
 PT New human interleukin-3 mutant polypeptide(s) - useful for treating various blood cell deficiencies, including leukopenia, neutropenia and leukaemia
 PS Claim 27: Page 99; 388pp; English.
 CC The sequence is a mutein of hIL-3, comprising Met-Ala-(15-125)hIL-3. The region 15-125 comprises the substitutions: 181, 25H, 29R, 32N, 37P, 42S, 45W, 51R, 55T, 59L, 62V, 67H, 69E, 73G, 76A, 79R, 82Q, 87S, 93S, 98T, 101A, 105Q, 109P, 116V, 120Q, 123E.
 CC The muteins may have biological activities similar to or better than hIL-3 and, in some cases, may also have an improved side effect profile, e.g. better therapeutic index. They may be used to treat e.g. leukopenia, neutropenia, thrombocytopenia, (aplastic) anemia, Chediak-Higashi syndrome, SLE, leukaemia, myelodysplastic syndrome, myelofibrosis; bone marrow suppression or haematopoietic deficiency associated with treatment with AZT or due to dialysis; or immunodeficiency resulting from viral infection, exposure to radiation or cancer treatment.
 CC The muteins are made by recombinant DNA techniques. Recombinant DNA sequences encoding the muteins are given in GENESEQ Accession Numbers
 CC Q68083 - Q68189.
 SO Sequence 113 AA:
 SQ 10 A; 7 R; 9 N; 5 D; 0 B; 2 C; 6 Q; 9 E; 0 Z; 2 G; 4 H; 10 I; 15 L; 4 K; 4 M; 3 F; 9 P; 5 S; 4 T; 1 W; 1 Y; 3 V;
 RS6067 Length: 113 April 28, 1998 12:48 Type: P Check: 6774 ..

1 MANCSIMIDE ITHHKRPPN PLIDRRNLNS EDMOIMERN LRTPNLLAFV
 51 RAVKHLNENAS GIEALIRNLQ PCLPSATAP SRHPIIKAG DMQEFREKLT
 101 FYLVTEQAO EQQ

11AA_SEQUENCE 1.0
 ID R56083 standard; peptide; 113 AA.
 AC R56083;
 DE 29-JUN-1995 (first entry)
 KW Human interleukin-3 mutein (pMon13402).
 OS Synthetic.
 PN W09412638-A.
 PD 09-JUN-1994.
 PF 22-NOV-1993: U11197.
 PR 24-NOV-1992: US-981044.
 PA (MONS) MONSANTO CO.
 PI (SEAR) SEARLE & CO G D.
 PI Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH, Easton AM, Klein BK, McKeam JP, Olins PO, Paik K; Thomas JW; WPI: 94-200266/24.
 DR N-PSDB; Q68169.
 PT New human interleukin-3 mutant polypeptide(s) - useful for treating various blood cell deficiencies, including leukopenia, neutropenia and leukaemia
 PS Claim 27: Page 130; 388pp; English.
 CC The sequence is a mutein of hIL-3, comprising Met-Ala-(15-125)hIL-3. The muteins may have biological activities similar to or better than hIL-3 and, in some cases, may also have an improved side effect profile, e.g. better therapeutic index. They may be used to treat e.g. leukopenia, neutropenia, thrombocytopenia, (aplastic) anemia,

CC Chediak-Higashi syndrome, SLE, leukaemia, myelodysplastic syndrome, myelofibrosis; bone marrow suppression or haematopoietic deficiency associated with treatment with AZT or due to dialysis; or immunodeficiency resulting from viral infection, exposure to radiation or cancer treatment. The muteins are made by recombinant DNA techniques. Recombinant DNA sequences encoding the muteins are given in GENESEQ Accession Numbers Q68083 - Q68189.
 SO Sequence 113 AA:
 SQ 11 A; 6 R; 8 N; 4 D; 0 B; 2 C; 6 Q; 9 E; 0 Z; 2 G; 4 H; 10 I; 15 L; 4 K; 3 M; 3 F; 8 P; 8 S; 4 T; 1 W; 1 Y; 4 V;
 RS6083 Length: 113 April 28, 1998 12:48 Type: P Check: 7352 ..

1 MANCSIMIDE LIHHLKIPN PSLDSANLNS EDVSIIMERN LRTPNLLAFV
 51 RAVKHLNENAS GIEALIRNLQ PCLPSATAP SRHPIIKAG DMQEFREKLT
 101 FYLVTEQAO EQQ

11AA_SEQUENCE 1.0
 ID R56136 standard; peptide; 113 AA.
 AC R56136;
 DE 29-JUN-1995 (first entry)
 KW Human interleukin-3 mutein (pMon13400).
 OS Synthetic.
 PN W09412638-A.
 PD 09-JUN-1994.
 PF 22-NOV-1993: U11197.
 PR 24-NOV-1992: US-981044.
 PA (MONS) MONSANTO CO.
 PI (SEAR) SEARLE & CO G D.
 PI Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH, Easton AM, Klein BK, McKeam JP, Olins PO, Paik K; Thomas JW; WPI: 94-200266/24.
 DR N-PSDB; Q68168.
 PT New human interleukin-3 mutant polypeptide(s) - useful for treating various blood cell deficiencies, including leukopenia, neutropenia and leukaemia
 PS Claim 32: Page 129-130; 388pp; English.
 CC The sequence is a mutein of hIL-3, comprising Met-Ala-(15-125)hIL-3. The muteins may have biological activities similar to or better than hIL-3 and, in some cases, may also have an improved side effect profile, e.g. better therapeutic index. They may be used to treat e.g. leukopenia, neutropenia, thrombocytopenia, (aplastic) anemia, Chediak-Higashi syndrome, SLE, leukaemia, myelodysplastic syndrome, myelofibrosis; bone marrow suppression or haematopoietic deficiency associated with treatment with AZT or due to dialysis; or immunodeficiency resulting from viral infection, exposure to radiation or cancer treatment. The muteins are made by recombinant DNA techniques. Recombinant DNA sequences encoding the muteins are given in GENESEQ Accession Numbers Q68083 - Q68189.
 SO Sequence 113 AA:
 SQ 12 A; 6 R; 8 N; 4 D; 0 B; 2 C; 6 Q; 9 E; 0 Z; 2 G; 4 H; 9 I; 14 L; 4 K; 3 M; 3 F; 9 P; 8 S; 4 T; 1 W; 1 Y; 4 V;
 RS6136 Length: 113 April 28, 1998 12:48 Type: P Check: 7287 ..

1 MANCSIMIDE AIHHLKIPN PSLDSANLNS EDVSIIMERN LRTPNLLAFV
 51 RAVKHLNENAS GIEALIRNLQ PCLPSATAP SRHPIIKAG DMQEFREKLT
 101 FYLVTEQAO EQQ

11AA_SEQUENCE 1.0
 ID R56115 standard; peptide; 113 AA.
 AC R56115;
 DE 26-JUN-1995 (first entry)
 KW Human interleukin-3 mutein (pMon13377).
 OS Synthetic.
 PN W09412638-A.

PD 09-JUN-1994.
PF 22-NOV-1993; U11197.
PR 24-NOV-1992; US-981044.
PA (MONS) MORSANTO CO.
PA (SEAR) SEARLE & CO G D.
PI Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH;
PI Thomas AM, Klein BK, McKearn JP, Olin PO, Paik K;
DR WPI; 94-200266/24.
DR N-PSDB; 068141.
PT New human interleukin-3 mutant polypeptide(s) - useful for
PT treating various blood cell deficiencies, including leukopenia,
PT neutropenia and leukaemia
PS Claim 32; Page 116-117; 388pp; English.
CC The sequence is a mutein of hIL-3, comprising Met-Ala-(15-125)hIL-3.
CC The muteins may have biological activities similar to or better than
CC hIL-3 and, in some cases, may also have an improved side effect
CC profile, e.g. better therapeutic index. They may be used to treat
CC e.g. leukopenia, neutropenia, thrombocytopenia, (aplastic) anemia,
CC Chediak-Higashi syndrome, SLE, leukaemia, myelodysplastic syndrome,
CC myelofibrosis; bone marrow suppression or haematopoietic deficiency
CC associated with treatment with AZT or due to dialysis; or
CC immunodeficiency resulting from viral infection, exposure to
CC radiation or cancer treatment. The muteins are made by recombinant
CC DNA techniques. Recombinant DNA sequences encoding the muteins are
CC given in GENESEQ Accession Numbers 068083 - 068189.
SQ Sequence 113 AA;
SQ 11 A; 7 R; 9 N; 6 D; 0 B; 2 C; 7 Q; 9 E; 0 Z; 2 G; 3 H;
SQ 9 I; 15 L; 4 K; 3 M; 3 F; 9 P; 6 S; 2 T; 1 W; 1 Y; 4 V;
R56115 Length: 113 April 28, 1998 12:48 Type: P Check: 7137 ..
1 MANCSIMIDE AIHHLKRPFA PLDDPNLND EDVSLMDRN LRLPLNESVY
51 RAVKLEENAS GIEALIRNLQ PCLPSATRAP SRAPITIKAG DMQERREKIQ
101 FYLVTEQAO EQQ
11AA_SEQUENCE 1.0
ID R10727 standard; Protein: 145 AA.
AC R10727.
DE 22-APR-1991 (first entry)
DE (Glu108Asp109Glu110)IL-3.
KW Interleukin-3; mutant; antagonist; cytopenia; immunodeficiency;
KW immunosuppression; charge reversal.
FH Key Location/Qualifiers
FT Peptide 1..12
FT /label- signal sequence
FT Mutation 120..122
FT /label- Arg-Arg-Lys -> Glu-Asp-Glu
PN EP-413383-A.
PD 20-FEB-1991.
PF 02-AUG-1990; 202117.
PF 14-AUG-1989; EP-202082.
PR 14-SEP-1989; EP-202331.
PR 02-AUG-1990; EP-202117.
PA (KONN) GIST-BROCADES NV.
PI Dorssers IJ, Vanleen RW;
DR WPI; 91-052685/08.
PT New interleukin-3 mutants - having deletions covering complete
PT coding sequence while retaining biological activity, have
PT antagonistic effect.
PS Claim 6; Page 18; 26pp; English.
CC The mutant was prepd. by recombinant DNA techniques using mutagenic
CC primers to change the codons of cDNA encoding IL-3. The peptide
CC pref. also has a deletion at the N-terminus (Aas 1-14) and/or the
CC C-terminus (Aas 116-133,120-130 and 130-133). The peptide has
CC antagonistic activity and can be used for therapeutic and diagnostic
CC purposes.
CC See also R10711-R10731 and R11051-R11055.
SQ Sequence 145 AA;
SQ 11 A; 5 R; 14 N; 8 D; 0 B; 2 C; 7 Q; 9 E; 0 Z; 3 G; 3 H;
SQ 10 I; 20 L; 6 K; 5 M; 5 F; 10 P; 8 S; 13 T; 2 W; 1 Y; 3 V;

R10727 Length: 145 April 28, 1998 12:48 Type: P Check: 7418 ..
1 MTMINSRBS VDAPIQTQTP LKTSWNCN MDEITTHLK QPPLPLDFN
51 NINGEDDIL MENNLRPNTL EAFNRVAKSL QNASALESTL KNLPLCPDLA
101 TAAPTRHPFH IKDGDWNEFE DELTFYIKTL ENAQAQOTTL SLAIF
11AA_SEQUENCE 1.0
ID R14839 standard; Protein: 464 AA.
AC R14839;
DE 30-JAN-1992 (first entry)
DE Protein deduced from human Glioblastoma cell cDNA.
KW RAS; oncogene; cancer; CAMP; p265.
OS Homo sapiens.
PN WO9116457-A.
PD 31-OCT-1991.
PF 19-APR-1991; U02714.
PR 20-APR-1990; US-511715.
PA (COLD-) COLD SPRING HARBOR.
PI Wiegler MH, Colicelli JT;
DR WPI; 91-339841/46.
DR N-PSDB; 014627.
PT Complementary screening for genes and prods. - e.g. RAS protein
PT and CAMP, that modify, complement or suppress genetic defect and
PT correct associated phenotypic alteration
PS Example 1B; Page 77; 169pp; English.
CC In the specification the sequence from which this amino acid
CC sequence was deduced is given the SEQ ID NO. 15 and
CC is described as the cDNA insert of plasmid p265. The insert
CC sequence shows no significant homology to previously isolated
CC genes. Plasmid p265 in E.coli (ATCC 68598) is described as
CC containing a human glioblastoma cell cDNA insert encoding a
CC RAS-related polypeptide. N.B. Many of the references in the text of
CC the specification to particular sequences are incorrect, therefore
CC caution is advised in accepting that this sequence is indeed p265 !
SQ Sequence 464 AA;
SQ 18 A; 22 R; 14 N; 28 D; 0 B; 9 C; 27 Q; 38 E; 0 Z; 17 G; 8 H;
SQ 20 I; 47 L; 34 K; 20 M; 22 F; 22 P; 39 S; 27 T; 3 W; 19 Y; 30 V;
R14839 Length: 464 April 28, 1998 12:48 Type: P Check: 4420 ..
1 MSISRSSSDS LEFDRSMPLF GYEADNSSL EYDEGSODE TMAPRIRKXK
51 KRSSSFVLPK LVKSOLQKVS GVFSSEMTPE KMWRRMAEL SRDKCTYFGC
101 LVQDYVSFLQ ENKECHVST DMLQTIQFWM TQVKNYLSQS SELDPIESL
151 IPEDQIDVTL EKAMHKCLIK PLKGAVEAML KQFNHAGSM KQKKNLQLY
201 KQNPQELGV FAPTRDFVDV EKIKVKFTM QKMSPEKKV MLLLRVCKLI
251 YTVMENNSGR MYGADDELPLV LTFVIAQCDW LELDTEIEYM MELDPSILH
301 GEGGYLLISA YGALSILKNF QEQQAARILS SETROTIRKM HKRRTNTTI
351 PSYDQFQNTL RVAFQVNSG CTGKTLIVRP YTTEDVYCOI CAEKFKYADP
401 EYSLFLFVN ETWOOLAEDT YPOKIRAEHL SRPQHIHF YKKRIKNDPY
451 GIIFQNGEED LTTS
11AA_SEQUENCE 1.0
ID R78367 standard; Protein: 125 AA.
AC R78367;
DE 27-JUN-1996 (first entry)
DE Human IL-3 mutant for multi-lineage haematopoietic cell prodn.
KW Interleukin-3; mutein; colony stimulating factor; CSF; chemotherapy;
KW radiation; haematopoiesis; IL-3 antagonist; antigenic; antibody;
KW synergy.
OS Synthetic.
PN WO9520977-A1.


```

CC The sequence is a mutein of hIL-3, comprising Met-Ala-(15-125)hIL-3.
CC The muteins may have biological activities similar to or better than
CC hIL-3 and, in some cases, may also have an improved side effect
CC profile, e.g. better therapeutic index. They may be used to treat
CC e.g. leukopenia, neutropenia, thrombocytopenia, (aplastic) anemia,
CC Chediak-Higashi syndrome, SLE, leukaemia, myelodysplastic syndrome,
CC myelofibrosis; bone marrow suppression or haematopoietic deficiency
CC associated with treatment with AZT or due to dialysis; or
CC immunodeficiency resulting from viral infection, exposure to
CC radiation or cancer treatment. The muteins are made by recombinant
CC DNA techniques. Recombinant DNA sequences encoding the muteins are
CC given in GENSEQ Accession Numbers Q68083 - Q68189.
SQ Sequence 113 AA:
SQ 11 A; 7 R; 9 N; 5 D; 0 B; 2 C; 6 Q; 9 E; 0 Z; 2 G; 4 H;
SQ 10 I; 15 L; 4 K; 2 M; 3 F; 9 P; 5 S; 4 T; 1 W; 1 Y; 4 V;
R56093 Length: 113 April 28, 1998 12:48 Type: P Check: 6987 ..

1 MANCSIAIDE ITHLKRPPN PLDPNNLNS EDVDILMERN LRTPNILAFV
51 RAVKHELNAS GIEALRNMQ PCLPSTAAP SRHPIIKAG DWQEFREKLT
101 FYLVTLQAO EQQ

!!AA_SEQUENCE 1.0
ID R56068 standard; peptide: 113 AA.
AC R56068:
DT 22-JUN-1995 (first entry)
DE Human interleukin-3 mutein (pMon13289).
KW Human interleukin-3; hIL-3; mutant; mutein.
OS Synthetic.
PN MO9412638-A.
PD 09-JUN-1994.
PF 22-NOV-1992; US-981044.
PR 24-NOV-1992; US-981044.
PA (MONS ) MONSANTO CO.
PA (SEAR ) SEARLE & CO G D.
PI Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH;
PI Easton AM, Klein BK, McKearn JP, Olins PO, Paik K;
PI Thomas JW;
PI WPI: 94-200266/24.
DR N-PSDB: Q68089.
PT New human interleukin-3 mutant polypeptide(s) - useful for
PT treating various blood cell deficiencies, including leukopenia,
PT neutropenia and leukaemia
PS Claim 27, Page 99; 388pp; English.
CC The sequence is a mutein of hIL-3, comprising Met-Ala-(15-125)hIL-3.
CC The region 15-125 comprises the substitutions: 181, 25H, 29V,
CC 32A, 37S, 42S, 45M, 51R, 55L, 59L, 62V, 67N, 69E, 73G, 76A, 79R,
CC 82Q, 87S, 93S, 98I, 101A, 105Q, 109E, 116V, 120Q, 123E.
CC The muteins may have biological activities similar to or
CC better than hIL-3 and, in some cases, may also have an improved side
CC effect profile, e.g. better therapeutic index. They may be used to
CC treat e.g. leukopenia, neutropenia, thrombocytopenia, (aplastic) anemia,
CC Chediak-Higashi syndrome, SLE, leukaemia, myelodysplastic syndrome,
CC myelofibrosis; bone marrow suppression or haematopoietic deficiency
CC associated with treatment with AZT or due to dialysis; or
CC immunodeficiency resulting from viral infection, exposure to radiation
CC or cancer treatment.
CC The muteins are made by recombinant DNA techniques. Recombinant DNA
CC sequences encoding the muteins are given in GENSEQ Accession Numbers
CC Q68083 - Q68189.
SQ Sequence 113 AA:
SQ 11 A; 6 R; 9 N; 5 D; 0 B; 2 C; 6 Q; 9 E; 0 Z; 2 G; 3 H;
SQ 10 I; 16 L; 4 K; 4 M; 3 F; 8 P; 6 S; 3 T; 1 W; 1 Y; 4 V;
R56068 Length: 113 April 28, 1998 12:48 Type: P Check: 6643 ..

1 MANCSIMIDE ITHLKVPPA PLDSNNLNS EDDMILMERN LRTPLILAFV
51 RAVKHELNAS GIEALRNMQ PCLPSTAAP SRHPIIKAG DWQEFREKLT
101 FYLVTLQAO EQQ

```

```

!!AA_SEQUENCE 1.0
ID R56092 standard; peptide: 113 AA.
AC R56092:
DT 26-JUN-1995 (first entry)
DE Human interleukin-3 mutein (pMon13407).
KW Human interleukin-3; hIL-3; mutant; mutein.
OS Synthetic.
PN MO9412638-A.
PD 09-JUN-1994.
PF 22-NOV-1993; U11197.
PF 22-NOV-1992; US-981044.
PR 24-NOV-1992; US-981044.
PA (MONS ) MONSANTO CO.
PA (SEAR ) SEARLE & CO G D.
PI Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH;
PI Easton AM, Klein BK, McKearn JP, Olins PO, Paik K;
PI Thomas JW;
PI WPI: 94-200266/24.
DR N-PSDB: Q68118.
PT New human interleukin-3 mutant polypeptide(s) - useful for
PT treating various blood cell deficiencies, including leukopenia,
PT neutropenia and leukaemia
PS Claim 32, Page 106; 388pp; English.
CC The sequence is a mutein of hIL-3, comprising Met-Ala-(15-125)hIL-3.
CC The muteins may have biological activities similar to or better than
CC hIL-3 and, in some cases, may also have an improved side effect
CC profile, e.g. better therapeutic index. They may be used to treat
CC e.g. leukopenia, neutropenia, thrombocytopenia, (aplastic) anemia,
CC Chediak-Higashi syndrome, SLE, leukaemia, myelodysplastic syndrome,
CC myelofibrosis; bone marrow suppression or haematopoietic deficiency
CC associated with treatment with AZT or due to dialysis; or
CC immunodeficiency resulting from viral infection, exposure to
CC radiation or cancer treatment. The muteins are made by recombinant
CC DNA techniques. Recombinant DNA sequences encoding the muteins are
CC given in GENSEQ Accession Numbers Q68083 - Q68189.
SQ Sequence 113 AA:
SQ 10 A; 7 R; 9 N; 5 D; 0 B; 2 C; 6 Q; 9 E; 0 Z; 2 G; 4 H;
SQ 10 I; 15 L; 4 K; 3 M; 3 F; 9 P; 5 S; 4 T; 1 W; 1 Y; 4 V;
R56092 Length: 113 April 28, 1998 12:48 Type: P Check: 7071 ..

1 MANCSIMIDE ITHLKRPPN PLDPNNLNS EDVDILMERN LRTPNILAFV
51 RAVKHELNAS GIEALRNMQ PCLPSTAAP SRHPIIKAG DWQEFREKLT
101 FYLVTLQAO EQQ

!!AA_SEQUENCE 1.0
ID R56101 standard; peptide: 113 AA.
AC R56101:
DT 26-JUN-1995 (first entry)
DE Human interleukin-3 mutein (pMon13425).
KW Human interleukin-3; hIL-3; mutant; mutein.
OS Synthetic.
PN MO9412638-A.
PD 09-JUN-1994.
PF 22-NOV-1993; U11197.
PF 24-NOV-1992; US-981044.
PA (MONS ) MONSANTO CO.
PA (SEAR ) SEARLE & CO G D.
PI Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH;
PI Easton AM, Klein BK, McKearn JP, Olins PO, Paik K;
PI Thomas JW;
PI WPI: 94-200266/24.
DR N-PSDB: Q68127.
PT New human interleukin-3 mutant polypeptide(s) - useful for
PT treating various blood cell deficiencies, including leukopenia,
PT neutropenia and leukaemia
PS Claim 32, Page 110; 388pp; English.
CC The sequence is a mutein of hIL-3, comprising Met-Ala-(15-125)hIL-3.
CC The muteins may have biological activities similar to or better than
CC hIL-3 and, in some cases, may also have an improved side effect
CC profile, e.g. better therapeutic index. They may be used to treat

```

CC e.g. leukopenia, neutropenia, thrombocytopenia, (aplastic) anemia,
 CC Chediak-Higashi syndrome, SLE, leukaemia, myelodysplastic syndrome,
 CC myelofibrosis; bone marrow suppression or haematopoietic deficiency
 CC associated with treatment with AZT or due to dialysis; or to
 CC immunodeficiency resulting from viral infection, exposure to
 CC radiation or cancer treatment. The mutants are made by recombinant
 CC DNA techniques. Recombinant DNA sequences encoding the mutants are
 CC given in GENESQ Accession Numbers Q68083 - Q68189.

SO Sequence 113 AA;
 SO 10 A: 7 R; 9 N; 5 D; 0 B; 2 C; 6 Q; 9 E; 0 Z; 2 G; 4 H;
 SO 11 I; 16 L; 4 K; 1 M; 3 F; 9 P; 5 S; 4 T; 1 W; 1 Y; 4 V;
 R56101 Length: 113 April 28, 1998 12:48 Type: P Check: 7006 ..

1 MANCISIIDE IHHKRPNN PLDPNNLNS EDVDILLERN LRTPNLLAFV
 51 RAVKHLLENAS GIEALIRNLQ PCLPSATAP SRHPITIKAG DMQEFREKLT
 101 FYLVTEQAQ EQQ

11AA_SEQUENCE 1.0
 ID R78353 standard; protein; 113 AA.
 AC R78353;
 DT 27-JUN-1996 (first entry)
 DE Human IL-3 mutant for multi-lineage haematopoietic cell prodn.
 KW Interleukin-3; mutetin; colony stimulating factor; CSF; chemotherapy;
 KW radiation; haematopoiesis; IL-3 antagonist; antigenic; antibody;
 KM Synthetic.
 OS WO9520977-A1.
 PN 10-AUG-1995.
 PD 02-FEB-1995; U01184.
 PE 04-FEB-1994; US-193373.
 PR (SEAR) SEARLE & CO G D.
 PA Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH;
 PI Easton AM, Klein BK, McKearn JP, Ollins PO, Palk K;
 PI Thomas JW;
 DR WPI; 95-283609/37.
 PT Multi-lineage hematopoietic cell production - by co-administration
 PT of human interleukin-3 variants and colony stimulating factors
 PT useful in radiation treatment and chemotherapy

PS Claim 11; Page 254; 317pp; English.
 CC R7837-R7838 are human interleukin-3 (IL-3) mutant proteins
 CC (mutetins). The IL-3 mutetins are used in a method for increasing the
 CC prodn. of multi-lineage haematopoietic cells in a mammal. The IL-3
 CC mutetins are co-administered with a colony stimulating factor
 CC selected from GM-CSF, CSF-1, G-CSF, Meg-CSF, erythropoietin, IL-1,
 CC IL-2, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13,
 CC LIF, flt3/lt2, human growth hormone, B-cell growth factor,
 CC B-cell differentiation factor, eosinophil differentiation factor
 CC and stem cell factor. Co-administration enhances therapeutic value
 CC through synergistic effects. The mutant IL-3 molecules can be used to
 CC treat conditions in which haematopoietic cells have been depleted
 CC due to e.g. chemotherapy or radiation treatment. A further use of
 CC the IL-3 variants is as antagonists of IL-3 or antigenic fragments
 CC for the prodn. of antibodies.

SO Sequence 113 AA;
 SO 11 A: 6 R; 9 N; 5 D; 0 B; 2 C; 6 Q; 9 E; 0 Z; 2 G; 3 H;
 SO 10 I; 16 L; 4 K; 4 M; 3 F; 8 P; 6 S; 3 T; 1 W; 1 Y; 4 V;
 R78353 Length: 113 April 28, 1998 12:48 Type: P Check: 6643 ..

1 MANCISIMIDE IHHKVPRA PLDDNNLNS EDMDLERN LRPNNLAFV
 51 RAVKHLLENAS GIEALIRNLQ PCLPSATAP SRHPITIKAG DMQEFREKLT
 101 FYLVTEQAQ EQQ

11AA_SEQUENCE 1.0
 ID R56100 standard; peptide; 113 AA.
 AC R56100;
 DT 26-JUN-1995 (first entry)
 DE Human interleukin-3 mutetin (pMon13424).

KW Human interleukin-3; hIL-3; mutant; mutetin.
 OS Synthetic.
 PN WO9412638-A.
 PD 09-JUN-1994.
 PE 22-NOV-1993; U11197.
 PR 24-NOV-1992; US-981044.
 PA (MONS) MONSANTO CO.
 PA (SEAR) SEARLE & CO G D.
 PI Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH;
 PI Easton AM, Klein BK, McKearn JP, Ollins PO, Palk K;
 PI Thomas JW;
 DR WPI; 94-200266/24.
 DR N-PSDB; Q68126.
 PT New human interleukin-3 mutant polypeptide(s) - useful for
 PT treating various blood cell deficiencies, including leukopenia,
 PT neutropenia and leukaemia

PS Claim 32; Page 110; 388pp; English.
 CC The sequence is a mutetin of hIL-3, comprising Met-Ala-(15-125)hIL-3.
 CC The mutants may have biological activities similar to or better than
 CC hIL-3 and, in some cases, may also have an improved side effect
 CC profile, e.g. better therapeutic index. They may be used to treat
 CC e.g. leukopenia, neutropenia, thrombocytopenia, (aplastic) anemia,
 CC Chediak-Higashi syndrome, SLE, leukaemia, myelodysplastic syndrome,
 CC myelofibrosis; bone marrow suppression or haematopoietic deficiency
 CC associated with treatment with AZT or due to dialysis; or
 CC immunodeficiency resulting from viral infection, exposure to
 CC radiation or cancer treatment. The mutetins are made by recombinant

11AA_SEQUENCE 1.0
 ID R56104 standard; peptide; 113 AA.
 AC R56104;
 DT 26-JUN-1995 (first entry)
 DE Human interleukin-3 mutetin (pMon13368).
 KW Human interleukin-3; hIL-3; mutant; mutetin.
 KM Synthetic.
 OS WO9412638-A.
 PN 09-JUN-1994.
 PD 22-NOV-1993; U11197.
 PE 24-NOV-1992; US-981044.
 PR (MONS) MONSANTO CO.
 PA (SEAR) SEARLE & CO G D.
 PI Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH;
 PI Easton AM, Klein BK, McKearn JP, Ollins PO, Palk K;
 PI Thomas JW;
 DR WPI; 94-200266/24.
 DR N-PSDB; Q68130.
 PT New human interleukin-3 mutant polypeptide(s) - useful for
 PT treating various blood cell deficiencies, including leukopenia,
 PT neutropenia and leukaemia

PS Claim 32; Page 111-112; 388pp; English.
 CC The sequence is a mutetin of hIL-3, comprising Met-Ala-(15-125)hIL-3.
 CC The mutants may have biological activities similar to or better than
 CC hIL-3 and, in some cases, may also have an improved side effect
 CC profile, e.g. better therapeutic index. They may be used to treat
 CC e.g. leukopenia, neutropenia, thrombocytopenia, (aplastic) anemia,
 CC Chediak-Higashi syndrome, SLE, leukaemia, myelodysplastic syndrome,
 CC myelofibrosis; bone marrow suppression or haematopoietic deficiency
 CC associated with treatment with AZT or due to dialysis; or
 CC immunodeficiency resulting from viral infection, exposure to
 CC radiation or cancer treatment. The mutetins are made by recombinant

1 MANCISIAIDE IHHKRPNN PLDPNNLNS EDVDILLERN LRTPNLLAFV
 51 RAVKHLLENAS GIEALIRNLQ PCLPSATAP SRHPITIKAG DMQEFREKLT
 101 FYLVTEQAQ EQQ

R56100 Length: 113 April 28, 1998 12:48 Type: P Check: 6550 ..

CC DNA techniques. Recombinant DNA sequences encoding the mutants are
 CC given in GENSEQ Accession Numbers Q68083 - Q68189.
 SQ Sequence 113 AA;
 SQ 12 A; 6 R; 9 N; 5 D; 0 B; 2 C; 6 Q; 9 E; 0 Z; 2 G; 3 H;
 SQ 9 I; 16 L; 4 K; 4 M; 3 F; 8 P; 6 S; 3 T; 1 W; 1 Y; 4 V;
 R56104 Length: 113 April 28, 1998 12:48 Type: P Check: 6555 ..
 1 MANCISIMIDE AIHHLKVPFA PLDSNNLNS EDMILMERN LRLPNLAFV
 51 RAVKLENAS GIEALIRNLQ PCLPSATAAP SRHPIITKAG DMQEFREKLT
 101 FYLVLEQAO EQQ
 !!AA_SEQUENCE 1.0
 ID R56094 standard; peptide; 113 AA.
 AC R56094;
 DT 26-JUN-1995 (first entry)
 DE Human Interleukin-3 mutein (pMon13415).
 KW Human Interleukin-3; hIL-3; mutant; mutein.
 OS Synthetic.
 PN WO9412638-A.
 PD 09-JUN-1994.
 PF 22-NOV-1993; U11197.
 PR 24-NOV-1992; US-981044.
 PA (MONS) MONSANTO CO.
 PI (SEAR) SEARLE & CO G D.
 PI Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH;
 PI Easton AM, Klein BK, McKearn JP, Olins PO, Palk K;
 PI Thomas JW, WPI: 94-200266/24.
 DR N-PSDB; Q68120.
 PT New human interleukin-3 mutant polypeptide(s) - useful for
 PT treating various blood cell deficiencies, including leukopenia,
 PT neutropenia and leukaemia
 PS Claim 32: Page 107; 388pp. English.
 CC The sequence is a mutein of hIL-3, comprising Met-Ala-(15-125)hIL-3.
 CC The mutants may have biological activities similar to or better than
 CC hIL-3 and, in some cases, may also have an improved side effect
 CC profile, e.g. better therapeutic index. They may be used to treat
 CC e.g. leukopenia, neutropenia, thrombocytopenia, (aplastic) anemia,
 CC Chediak-Higashi syndrome, SLE, leukaemia, myelodysplastic syndrome,
 CC myelofibrosis; bone marrow suppression or haematopoietic deficiency
 CC associated with treatment with AZT or due to dialysis; or
 CC immunodeficiency resulting from viral infection, exposure to
 CC radiation or cancer treatment. The mutants are made by recombinant
 CC DNA techniques. Recombinant DNA sequences encoding the mutants are
 CC given in GENSEQ Accession Numbers Q68083 - Q68189.
 SQ Sequence 113 AA;
 SQ 10 A; 7 R; 9 N; 5 D; 0 B; 2 C; 6 Q; 9 E; 0 Z; 2 G; 4 H;
 SQ 11 I; 15 L; 4 K; 2 M; 3 F; 9 P; 5 S; 4 T; 1 W; 1 Y; 4 V;
 R56094 Length: 113 April 28, 1998 12:48 Type: P Check: 7043 ..
 1 MANCISIIDE IHHHLKRPFN PLDPNNLNS EDVILMERN LRLPNLAFV
 51 RAVKLENAS GIEALIRNLQ PCLPSATAAP SRHPIITKAG DMQEFREKLT
 101 FYLVLEQAO EQQ
 !!AA_SEQUENCE 1.0
 ID R22820 standard; Protein; 134 AA.
 AC R22820;
 DT 04-SEP-1992 (first entry)
 DE K64, P83, M88, K113, M127, E132, S134 hIL-36 derivative.
 KW Human, Interleukin-3; bone marrow transplant; graft; platelet.
 OS Homo sapiens.
 PN J04063595-A.
 PD 28-FEB-1992.
 PR 19-OCT-1990; JP-087468.
 PR 03-APR-1990; JP-087468.
 PR 19-OCT-1990; JP-279108.
 PA (KIRI) KIRIN BREWERY KK.

DR WPI: 92-120155/15.
 DR N-PSDB; Q22508.
 PT Human Interleukin 3 deriv. and its prepn. - for supplementing
 PT bone marrow transplantation and increasing platelet count
 PS Disclosure; Fig 8; 21pp; Japanese.
 CC This derivative of human IL-3 has amino acids Arg, Leu, Leu, Thr,
 CC Ala, Thr, Ala and Phe at positions 64, 83, 88, 113, 122, 127, 132
 CC and 134, respectively, substituted by Lys, Pro, Met, Lys, Glu, Met,
 CC Glu and Ser. The derivative has a higher activity than native
 CC IL-3. See R22813-4, Q22503-Q22510 and Q22513.
 SQ Sequence 134 AA;
 SQ 9 A; 5 R; 13 N; 6 D; 0 B; 2 C; 7 Q; 9 E; 0 Z; 2 G; 3 H;
 SQ 9 I; 18 L; 9 K; 6 M; 4 F; 10 P; 8 S; 9 G; 2 T; 2 W; 1 Y; 2 V;
 R22820 Length: 134 April 28, 1998 12:48 Type: P Check: 6016 ..
 1 MAPMTQTSL KISWVNCNM IDEITHLKO PPLPLDPFN LNEGDDILM
 51 ENNLRRPMLE AFNKAVALQ NASALSIK NLPCLPMAT AAPTRHPDHI
 101 KDGMNEFR KLFYKLTLE NEQAQMTLS LEIS
 !!AA_SEQUENCE 1.0
 ID R56139 standard; peptide; 113 AA.
 AC R56139;
 DT 29-JUN-1995 (first entry)
 DE Human Interleukin-3 mutein (pMon13403).
 KW Human Interleukin-3; hIL-3; mutant; mutein.
 OS Synthetic.
 PN WO9412638-A.
 PD 09-JUN-1994.
 PF 22-NOV-1993; U11197.
 PR 24-NOV-1992; US-981044.
 PA (MONS) MONSANTO CO.
 PI (SEAR) SEARLE & CO G D.
 PI Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH;
 PI Easton AM, Klein BK, McKearn JP, Olins PO, Palk K;
 PI Thomas JW, WPI: 94-200266/24.
 DR N-PSDB; Q68172.
 PT New human interleukin-3 mutant polypeptide(s) - useful for
 PT treating various blood cell deficiencies, including leukopenia,
 PT neutropenia and leukaemia
 PS Claim 32: Page 131; 388pp. English.
 CC The sequence is a mutein of hIL-3, comprising Met-Ala-(15-125)hIL-3.
 CC The mutants may have biological activities similar to or better than
 CC hIL-3 and, in some cases, may also have an improved side effect
 CC profile, e.g. better therapeutic index. They may be used to treat
 CC e.g. leukopenia, neutropenia, thrombocytopenia, (aplastic) anemia,
 CC Chediak-Higashi syndrome, SLE, leukaemia, myelodysplastic syndrome,
 CC myelofibrosis; bone marrow suppression or haematopoietic deficiency
 CC associated with treatment with AZT or due to dialysis; or
 CC immunodeficiency resulting from viral infection, exposure to
 CC radiation or cancer treatment. The mutants are made by recombinant
 CC DNA techniques. Recombinant DNA sequences encoding the mutants are
 CC given in GENSEQ Accession Numbers Q68083 - Q68189.
 SQ Sequence 113 AA;
 SQ 10 A; 5 R; 9 N; 6 D; 0 B; 2 C; 7 Q; 8 E; 0 Z; 2 G; 4 H;
 SQ 10 I; 15 L; 4 K; 4 M; 3 F; 10 P; 7 S; 4 T; 1 W; 1 Y; 1 V;
 R56139 Length: 113 April 28, 1998 12:48 Type: P Check: 6301 ..
 1 MANCISIMIDE IHHHLKRPFN PLDPNNLNS EDMILMDSN LRLPNLAFV
 51 HASQOLENAS GIEALIRNLQ PCLPSATAAP SRHPIITKAG DMQEFREKLT
 101 FYLVLEQAO EQQ
 !!AA_SEQUENCE 1.0
 ID W15514 standard; Protein; 115 AA.
 AC W15514;
 DT 28-NOV-1997 (first entry)
 DE Interleukin-3 receptor agonist pMON1155.pcp.

KM Interleukin-3 receptor agonist; stem cell, hematopoietic cell;
 OS human; gene therapy.
 OS Chimeric Homo sapiens;
 OS Chimeric synthetic.
 PN MO9712979-A1.
 PD 10-APR-1997.
 PF 04-OCT-1996; U15941.
 PR 05-OCT-1995; US-004835.
 PA (SEAR) SEARLE & CO G D.
 PI Bauer SC, Baum CM, Caparon MH, Feng Y, McKearn JP;
 PI McWhorter CA, Zurfluhll;
 DR N-PSDB; T60351.
 PT New peptide(s) with IL-3 receptor agonist activity - used to
 PT stimulate proliferation of haematopoietic cells, or for expansion of
 PT stem cells in vitro
 PS Claim 3; Page 11; 126pp; English.
 CC This polypeptide, designated pMON31155.pep, is a novel interleukin-3
 CC (IL-3) receptor agonist based on human IL-3 (see W14061) and
 CC prepared by multiple PCR of the IL-3 gene and DNA fragment
 CC rearrangement (see also T60351). Claimed IL-3 receptor agonists
 CC (W1514-21) are produced in host cells using a claimed method.
 CC They can be used to stimulate production of haematopoietic cells in
 CC a patient (claimed), e.g. one who has had chemotherapy or
 CC radiotherapy, or is affected by viral infection or some other
 CC immune deficiency. They can also be used for selective ex vivo
 CC expansion of stem cells (claimed) for return to a patient for
 CC treatment of haematopoietic disorders, optionally after insertion
 CC of DNA into the cultured cells. Compared with IL-3, the agonists
 CC may have improved activity (allowing a lower dose) and may also
 CC have reduced side effects and better physical properties such as
 CC solubility, stability and refold efficiency.
 SQ Sequence 115 AA:
 SQ 9 A; 7 R; 9 N; 6 D; 0 B; 2 C; 6 Q; 9 E; 0 Z; 5 G; 3 H;
 SQ 10 I; 15 L; 4 K; 2 M; 3 F; 9 P; 7 S; 3 T; 1 W; 1 Y; 4 V;
 W15514 Length: 115 April 28, 1998 12:48 Type: P Check: 8744 ..

1 LDPNNINDE VSIIDNRNR LPNLESEFVRA VKNLENASGI EAILRNIQPC
 51 LPSATAPSR HPIIKAGDM QEPREKLFY LVLEQAOQ QGGGSCS1M
 101 IDEIRHMKR PPAPL

!!AA_SEQUENCE 1.0
 ID R56141 standard; peptide; 106 AA.
 AC R56141;
 DT 29-JUN-1995 (first entry)
 DE Human Interleukin-3 mutein (pMON13412).
 KW Human Interleukin-3; hIL-3; mutant; mutein.
 OS Synthetic.
 PN WO9412638-A.
 PD 09-JUN-1994.
 PF 22-NOV-1993; U11197.
 PR 24-NOV-1992; US-961044.
 PA (MONS) MONSANTO CO.
 PA (SEAR) SEARLE & CO G D.
 PI Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH;
 PI Easton AM, Klein BK, McKearn JP, Olin PO, Paik K;
 PI Thomas JW;
 DR N-PSDB; Q68175.
 PT New human interleukin-3 mutant polypeptide(s) - useful for
 PT treating various blood cell deficiencies, including leukopenia,
 PT neutropenia and leukemia
 PS Claim 32; Page 132; 38pp; English.
 CC The sequence is a mutein of hIL-3, comprising Met-Ala-(15-116)hIL-3.
 CC The muteins may have biological activities similar to or better than
 CC hIL-3 and, in some cases, may also have an improved side effect
 CC profile, e.g. better therapeutic index. They may be used to treat
 CC e.g. leukopenia, neutropenia, thrombocytopenia, (aplastic) anemia,
 CC Chediak-Higashi syndrome, SLE, leukemia, myelodysplastic syndrome,
 CC myelofibrosis; bone marrow suppression or haematopoietic deficiency

CC associated with treatment with AZT or due to dialysis; or
 CC immunodeficiency resulting from viral infection, exposure to
 CC radiation or cancer treatment. The muteins are made by recombinant
 CC DNA techniques. Recombinant DNA sequences encoding the muteins are
 CC given in GENESQ Accession Numbers Q68083 - Q68189.
 SQ Sequence 106 AA:
 SQ 9 A; 7 R; 9 N; 5 D; 0 B; 2 C; 3 Q; 6 E; 0 Z; 2 G; 4 H;
 SQ 10 I; 16 L; 4 K; 4 M; 3 F; 9 P; 6 S; 3 T; 1 W; 1 Y; 2 V;
 R56141 Length: 106 April 28, 1998 12:48 Type: P Check: 8813 ..

1 MANCSINIDE ITHIKRPPN PLDPNNINS EDMIDIMERN LTPN1LAFV
 51 RAVKHLENAS GIEAIRNLQ PCLPSATAAP SRPIIKAG DWQEFRLKIQ
 101 FYLSTL

!!AA_SEQUENCE 1.0
 ID R36739 standard; Protein; 245 AA.
 AC R36739;
 DT 13-SEP-1993 (first entry)
 DE Human bone morphogenetic protein BMP-3.
 KW Induction; bone defect treatment; healing; wound; ulcers;
 KW injury; tissue repair; osteoporosis; burns; incisions;
 KW neuronal survival increase; fracture reduction; cartilage growth;
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Region 142
 FT /note "x represents 229 missing amino acids"
 PN WO9309229-B.
 PD 13-MAY-1993.
 PF 02-NOV-1992; U09430.
 PR 04-NOV-1991; US-787496.
 PR 07-APR-1992; US-864692.
 PA (GEMT) GENETICS INST INC.
 PI Israel D, Wolfman NM.
 DR N-PSDB; Q41298.
 PT Recombinant hetero-dimeric BMP proteins - are useful in treating
 PT bone defects, healing bone injury and in wound healing
 PS Disclosure; Page 117; 169pp; English.
 CC The sequence is that of the human bone morphogenetic protein
 CC BMP-3. The 229 missing amino acids represent a page missing from
 CC the sequence as given in the specification.
 SQ Sequence 245 AA:
 SQ 19 A; 17 R; 8 N; 8 D; 0 B; 10 C; 7 Q; 15 E; 0 Z; 18 G; 2 H;
 SQ 10 I; 24 L; 12 K; 5 M; 11 F; 19 P; 22 S; 10 T; 4 W; 8 Y; 15 V;
 1 Others;
 R36739 Length: 245 April 28, 1998 12:48 Type: P Check: 5590 ..

1 MAGASRLTL WLGCRCVSLA QGERKPPFP ELRKAVPGDR TAGGPDSEL
 51 QPODKVSEHM LRLYDRYSTV QAARTPGSLG GGSQPMRPRL LREGTVBSF
 101 RAAAEETLER KGLYFNLTLS LTKSENTLSA TLYFCGEIG NNCARYLKY
 151 DRADIGWSEW IISPKSEFAY YCSGACQFPM PKSLKASNNHA TIQSYRAVG
 201 VVFGIPEPC VPEKMSLSI LFEDENKNV LKYPNMIVE SCACR

!!SEQUENCE_LIST 1.0

(Nucleotide) FASTA of: seq1b from: 1 to: 597 April 28, 1998 08:41

TO: geneseqn:* Sequences: 159,646 Symbols: 57,748,920 Word Size: 6

Sequences too short to analyze: 10 (42 symbols)

Databases searched: Geneseq-NA, Release 30.0, Released on 26Jan1998, Formatted on 10Mar1998

Searching with both strands of the query.

Scoring matrix: GenRunData:fastadna.cmp

Constant pamfactor used

Gap creation penalty: 16 Gap extension penalty: 4

Histogram Key:

Each histogram symbol represents 814 search set sequences
Each inset symbol represents 4 search set sequences
z-scores computed from opt scores

```

z-score obs exp
(-) (*)
< 20 610 0 : *
22 38 0 : *
24 25 0 : *
26 38 6 : *
28 95 66 : *
30 205 400 : *
32 611 1546 : *
34 1710 4192 : *
36 3933 8609 : *
38 7111 14227 : *
40 11723 19846 : *
42 23395 24259 : *
44 33392 26760 : *
46 48828 27255 : *
48 36046 26094 : *
50 29756 23811 : *
52 25699 20934 : *
54 13874 17881 : *
56 11616 14936 : *
58 9440 12262 : *
60 6740 9933 : *
62 6220 7963 : *
64 4366 6333 : *
66 3549 5006 : *
68 2740 3937 : *
70 2015 3086 : *
72 1633 2411 : *
74 1221 1880 : *
76 839 1463 : *
78 697 1137 : *
80 586 883 : *
82 452 676 : *
84 333 535 : *
86 230 414 : *
88 188 320 : *
90 195 248 : *
92 153 192 : *
94 112 148 : *
96 76 115 : *
98 46 89 : *
100 38 69 : *
102 35 53 : *
104 27 41 : *
106 32 32 : *
108 30 25 : *
110 15 19 : *

```

```

112 18 15 : *
114 2 11 : *
116 0 9 : *
118 8 7 : *
>120 20 5 : *

```

Results sorted and z-values calculated from opt score
1988 scores saved that exceeded 81
125983 optimizations performed
Joining threshold: 52, optimization threshold: 37, opt. width: 16

The best scores are:

```

GENESQN:T49312 Begin: 1 End: 597 Init1 initn opt z-sc E(290156)..
! Human endothelial-monocyte activati... 2985 2985 2985 3056.5 0
GENESQN:086718 Begin: 506 End: 746
! MOUSE EMAP11 CDNA... New endothelia... 424 524 449 452.9 2.2e-18
GENESQN:094113 Begin: End: 389 Strand: -
! PML2 genomic DNA... Thrombopoietin ... 72 72 136 132.3 1.7
\\End of List

```

seq1b
GENESQN:T49312

```

ID T49312 standard; cDNA; 636 BP.
AC T49312;
DT 04-JUL-1997 (first entry)
DE Human endothelial-monocyte activating polypeptide III encoding cDNA.
KW Endothelial-monocyte activating polypeptide; EMAP III; human;
KW neoplasia; ss. . . .

```

SCORES Init1: 2985 Initn: 2985 Opt: 2985 z-score: 3056.5 E(): 0
100.0% identity in 597 bp overlap

```

seq1b 10 20 30 40 50 60
TACCCTGCGCTGAAAAAATCGAGCGTGCCTACCCAGATCCCTCAAGCAGAGCC
|||||
T49312 TACCCTGCGCTGAAAAAATCGAGCGTGCCTACCCAGATCCCTCAAGCAGAGCC
10 20 30 40 50 60

seq1b 70 80 90 100 110 120
AATGCCCAAGGCGCTCCCAAGATTCAAGACAGAGAGAGTATCCATCCCGCGTGAT
|||||
T49312 AATGCCCAAGGCGCTCCCAAGATTCAAGACAGAGAGAGTATCCATCCCGCGTGAT
70 80 90 100 110 120

seq1b 130 140 150 160 170 180
ATCCGTGTGGGGAATATCACTGTGAGAGACACCGACTGTGAGCGCGCTGATGTA
|||||
T49312 ATCCGTGTGGGGAATATCACTGTGAGAGACACCGACTGTGAGCGCGCTGATGTA
130 140 150 160 170 180

seq1b 190 200 210 220 230 240
GAGAGATTGACCTGGGGAAGCTGAACCGAGCTGTGAGCGCGCTGATGTA
|||||
T49312 GAGAGATTGACCTGGGGAAGCTGAACCGAGCTGTGAGCGCGCTGATGTA
190 200 210 220 230 240

seq1b 250 260 270 280 290 300
GTGCCCAAGGAGAACTGCAAGCAGCGCTGTGAGCGCGCTGATGTA
|||||
T49312 GTGCCCAAGGAGAACTGCAAGCAGCGCTGTGAGCGCGCTGATGTA
250 260 270 280 290 300

seq1b 310 320 330 340 350 360
AAGATGAGAGAGATCGAGTCCCAAGCAGCTTCTGTGCTTCTAATAGAGGATTAAC
|||||
T49312 AAGATGAGAGAGATCGAGTCCCAAGCAGCTTCTGTGCTTCTAATAGAGGATTAAC
310 320 330 340 350 360

seq1b 370 380 390 400 410 420
CGCCAGGTGAACCTCTGACCCCTCCGCGAGCGCTCTCTCTGTGAGCAGCGTGTGTC
|||||
T49312 CGCCAGGTGAACCTCTGACCCCTCCGCGAGCGCTCTCTCTGTGAGCAGCGTGTGTC
370 380 390 400 410 420

```

```

T49312      |||||||
CGCCAGGTTGAACCTTGGACCTCGGACGCTGCTCCGTCGAGCAGTGTGTTG
370          380          390          400          410          420

seq1b      |||||||
AAGGCTTTGAAAAGGCGCCACAGATGAGAGCTCAAGCCAGAAAGAGTCTTCGAG
430          440          450          460          470          480
T49312      |||||||
AAGGCTTTGAAAAGGCGCCACAGATGAGAGCTCAAGCCAGAAAGAGTCTTCGAG
430          440          450          460          470          480

seq1b      |||||||
AAGTTGCAGGCTGACTTCAAAATTTCTGAGAGTGCATCGCACAGTGGAAAGCAACAC
490          500          510          520          530          540
T49312      |||||||
AAGTTGCAGGCTGACTTCAAAATTTCTGAGAGTGCATCGCACAGTGGAAAGCAACAC
490          500          510          520          530          540

seq1b      |||||||
TTCATGACCAAGCTGGGCTCCATTTCCGTAAATGCTGAAAGGGGGAACATTAGC
550          560          570          580          590
T49312      |||||||
TTCATGACCAAGCTGGGCTCCATTTCCGTAAATGCTGAAAGGGGGAACATTAGCTAG
550          560          570          580          590          600

T49312      |||||||
CCAGCCGACGACTTTCCTCCCTTCTTCACCACTGA
610          620          630

seq1b      |||||||
GENESEQN:Q86718
ID Q86718 standard; cDNA; 1086 BP.
AC Q86718;
DE 29-SEP-1995 (first entry)
DE Mouse EMAP11 cDNA.
KW EMAP11; endothelial monocyte activating polypeptide II; chemotaxis;
inflammation; tissue factor; tumor; cancer; therapy; metha; ...

SCORES      Init1: 424 Initn: 524 Opt: 449 z-score: 452.9 E(): 2.2e-18
65.1% identity in 241 bp overlap

seq1b      |||||||
GGCTTGCACAGATTCAGAACCAAGAGAGTATCCATCCCGGCTGGATTCCTGTGG
80          90          100          110          120          130
Q86718      |||||||
CGCAGCAGCAGCAGTCTGACTCAAGCCATGACGACGATCGGCTGTGCAATTTG
480          490          500          510          520          530

seq1b      |||||||
GGAAATCATCATCTGTGGAGAGCACCCAGATGACAGACAGCTGTATGTAGAGATTTG
140          150          160          170          180          190
Q86718      |||||||
GTTGTATGTCTACTCTCCAGAGAGCACCCCTGATGCACTGTATGTGAGAGATAG
540          550          560          570          580          590

seq1b      |||||||
ACGTGGGGAGACTGACCTGAGCTGTGTGAGGGGCTGTACAGTTCTGCGCCAGAG
200          210          220          230          240          250
Q86718      |||||||
ATGTGGGAGAACACCGCCGCGCAGCGTGTGTCAGCGGGCTGTGATCATGTTCTCTAG
600          610          620          630          640          650

seq1b      |||||||
AGCAACTGACAGCAGAGCTGTAGTGTCTGTGTGCACTTGAACCCAGAAAGATGAGAG
260          270          280          290          300          310
Q86718      |||||||
AACAGATGCAAAATGTATGAGTGTGTATCTGTATCTGTAAAGCTGCAAAAGATGCGGG
660          670          680          690          700          710

seq1b      |||||||
GAGTCGAGTCCCAAGGAGCTTTCTGTGTGTCTTCTATTAAGGATTAACCCCAAGTTG
320          330          340          350          360          370
Q86718      |||||||
GAGTCTGTCTCAAGCAGATGTATGTGTGCAAGTTCACCAAGAGAAAGTGAATTTCTG
720          730          740          750          760          770

seq1b      |||||||
AACTCTTGAGACCTCCGCGAGGCTGTCTGTCTGTGAGAGCAGTGTGTAAGGCGCTAG
380          390          400          410          420          430
Q86718      |||||||
CCCTCTCCCAACGAGGCTCTCTCTGTGGGACAGATTAATTGATGCTTTCTGTGAGAGC

```

```

seq1b /rev
GENESEQN:Q94113
ID Q94113 standard; DNA; 1014 BP.
AC Q94113;
DE 27-FEB-1996 (first entry)
DE PM12 genomic DNA.
KW Human; thrombopoietin; TPO; mpl ligand; hML; fragment polypeptide;
megakaryocytopoietic cytokine receptor; thrombopoietic signal; ...

SCORES      Init1: 72 Initn: 72 Opt: 136 z-score: 132.3 E(): 1.7
60.5% identity in 152 bp overlap

seq1b      |||||||
CCTTCTATAGAGACACAGACAGATGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG
349          359          369          379          389          399
Q94113      |||||||
TGGAGCAGATGATGACAGCAGCGGGGAGACAGAGGAGGAGCCCTTGCTCTCATTC
220          230          240          250          260          270

seq1b      |||||||
TTTCAGGT--TGACAG--CACCAGTACCAAGCTGTCTGCAAGTCTCTTTGG--G
289          299          309          319          329          339
Q94113      |||||||
GTGCAGCTTCTGTGACAGGCTTCCTCCTCCTCGGCGCCGCGAGCAGCTCTGGAATG
280          290          300          310          320          330

seq1b      |||||||
CAGCAACTGATACAGGCGGCTGACCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTG
239          249          259          269          279          289
Q94113      |||||||
CAGGGA--AGACCA--CAGCTCACAGAGATCCAGTGCATCTTCTGCACTTCCACAA
340          350          360          370          380          390

seq1b      |||||||
CTTACATACAGGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
179          189          199          209          219          229
Q94113      |||||||
CTGCTCCGAGGAAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
400          410          420          430          440          450

! CPU time used:
! Database scan: 0:11:13.2
! Post-scan processing: 0:00:00.2
! Total CPU time: 0:11:13.5
! Output file: seq1b.fasta
! ID T49312 standard; cDNA; 636 BP.
! AC T49312;
! DE 04-JUL-1997 (first entry)
! DE Human endothelial-monocyte activating polypeptide III encoding cDNA.
! KW Endothelial-monocyte activating polypeptide; EMAP III; human;
! KW neoplasia; ss.
! OS Synthetic.
! FT Key Location/Qualifiers
! FT CDS 94..600
! FT /tag= a
! FT /product= EMAP-III
! FT W03640719-A1.
! PD 19-DEC-1996.
! PF 07-JUN-1995; U07328.
! PR 07-JUN-1995; W0-007328.
! PA (HUMA-) HUMAN GENOME SCI INC.
! PI Coleman TA, Olsen HS, Rosen CA;
! DR WPI: 97-052219/05.
! DR P-PSDB: W06596.
! PT New isolated endothelial-monocyte activating polypeptide III - used
! PT to develop prods. for the diagnosis and treatment of conditions
! PT involving EMAP III e.g. neoplasia
! PS Claim 5; Page 37; 50pp; English.
! CC The present sequence encodes the 168 amino acids of the novel
! CC polypeptide endothelial-monocyte activating polypeptide III (EMAP III).

```


CC The product can be used for the diagnosis and treatment of conditions
 CC involving abnormal expression of EMAP III. In particular, EMAP III can
 CC be used to regress neoplasia such as in tumours and cancers.
 CC EMAP III shows a high degree of homology to EMAP II with 60%
 CC identity and 75% similarity over a 150 amino acid stretch.
 CC Sequence 636 BP; 172 A; 166 C; 178 G; 120 T;

T49312 Length: 636 April 28, 1998 12:51 Type: N Check: 4493 ..

```

1  TACCCCTGCC CTGAAAAAC TGGCCACGCG TGCTTACCCA GATCCCTCAA
51  AGCAGAAAGCC AATGGCCCAA GGCCTGCCAA GAATTCAGAA CCAGAGGAGG
101  TCATCCCATC CCGGCTGGAT ATCCGTGTGG GGAATAATCAT CACTGTGGAG
151  AAGCACCAGG ATGCAGACAG CCGTATATGA GAGAAGATTG ACGTGGGGGA
201  AGCTAACCAC CGAAGCTGGG TGAGCGGCGT GGTACAGTTC GTGCCAAGG
251  AGGAAGTCCA GGACAGGCTG GTAAGTGTGC TGTCAACCTT GAAACCCGAG
301  AAGATGAGAG GAGTCGAGTC CCAAGGCATG CTTCTGTGTG CTTCTATAGA
351  AGGATTAAC CGCCAGGTTG AACCTTGA CCCTCCGGCA GCGCTGTGTC
401  CTGTGAGCA CGTGTGTGTG AAGGGCTATG AAAAGGGCCA ACCAGATGAG
451  GAGCTCAGC CCAAGAAAGA AGTCTTGAG AGTTGCAGG CTGACTTCAA
501  AATTCTGAG GAGTGCATCG CACAGTGA GCAAAACCAAC TTCATGACCA
551  AGCTGGGCTC CATTTCTGT AAATCCCTGA AAGGGGGGAA CATAGTAGAG
601  CCAGCCACAG ATCTTCCGCC CTCTTCCAC CACTGA

```

```

!!NA_SEQUENCE 1.0
ID 086718 standard; cDNA; 1086 BP.
AC 086718;
DT 29-SEP-1995 (first entry)
DE Mouse EMAPII cDNA.
KW EMAPII; endothelial monocyte activating polypeptide II; chemotaxis;
  inflammation; tissue factor; tumor; cancer; therapy; metha;
  sarcoma; ss.
OS Mus musculus.
FH Key Location/Qualifiers
FT CDS 64..996
FT /tag- a
FT W09509180-A.
FT 06-APR-1995.
FT 29-SEP-1994; U11085.
FT 29-SEP-1993; U1-129456.
PA (UYCO ) UNIV COLUMBIA NEW YORK.
PI Clauss M, Kao J, Kayton M, Libutti SK, Stern DM;
  WPI: 95-147389/19.
DR P-PSDB; R72577.
PT New endothelial monocyte activating polypeptide II - induces
  chemotaxis, inflammation and tissue factor, useful for treating
  tumors, also related antibodies, DNA and active fragments
  PS Disclosure; Fig.4; 180pp; English.
CC A mouse metha sarcoma cDNA library was screened with a probe
  based on the N-terminal sequence of mouse EMAPII. Overlapping
  clones were combined to obtain a contiguous full-length sequence
  CC (given in 086718) encoding a 33 kDa protein (R72577).
CC Recombinant EMAPII was expressed in E. coli.
CC Sequence 1086 BP; 346 A; 212 C; 297 G; 231 T;
SO

```

086718 Length: 1086 April 28, 1998 12:51 Type: N Check: 3316 ..

```

1  GAGGCTGCTC AAGAGCTCG GTGGGTCAC CGCTTATGT TTCTGTGCGG
51  ATTCTGGGGA AAGATGGCAA CGAATGATGC TGTCTGAAG AGGCTGGAGC
101  AGAAGGTTGC AGAGCGGAT CAGATCATCG AATATCTCAA GCAGCAGGTT

```

```

151  GCTCTTCTTA AGAGAAAGC AATTITGAG GCACATGA GAGAAGAAAA
201  GAAATTGGA GTTAAATG CTTAACTGAA AAAAAGATA GAAGAGCTAA
251  AGCAAGAGCT GATTCGTGCA GAAATTCANA ACGGAGTGA GCAAGTGCGT
301  GTTCATGTA GTACTCCACT GCAGAGAAC TGTACTGCTT CTGAAGAGTG
351  GGTGAGTCT CATTCACTAG CAACCAACCG CTCTCTGCTT ACNAAAGAGC
401  AGATCAAGC GGGAGAAAG AAGAGGTGA AAGAGAAGC TGAAGAGAAA
451  GGAGAGAAA AGAGAAAGCA GCAGTGGCA GCAGCAAGTA CTGACTCCAA
501  GCTATCAGC GCATCGGCTC TGGATCTTG AATTGTTGT ATTTTACTG
551  CCAGAAGCA CCTGATGCA GATTCACTGT ATGTGAGGA AGTAGATGTG
601  GGAGAAGCAG CCGCGCGAC GGTGCTCAG GGGCTGTGTA ATCATGTTCC
651  TCTAAGACAG ATGCAAAATC GTATGTTGT TTTACTCTGT AATCTGAAGC
701  CTGCAAGAT GCGGGGAGT CTGTCTCAAG CCATGTTGAT GTGTGCCAGT
751  TCACCAAGA AAGTGAAGT TGTGGCCCTT CCCAAGGGGT CCGTTCCTGG
801  GGACGAATT ACTTTTATG CTTTCTCTGG AGAGCTGAC AAGAGCTTAA
851  ACCCTAAGA GAAGATCTGG GAGCAGATCC AGCCTGACT GCACACCAAT
901  GGTGAGTGG TGCCACATA CAAAGAGCT CCCTTTGAG TGAAGGGGAA
951  GGGATTGTC AGAGCCCAA CCAATGCCAA TAGTGAATT AAATAGTGC
1001  TCTGTACTG AAGACATATG GCGAAACTT AATAACAATA AAGAGAAGTG
1051  TGTATATCAC TTACATATTA AAAAAAAAAA AAAAAA

```

```

!!NA_SEQUENCE 1.0
ID 094113 standard; DNA; 1014 BP.
AC 094113;
DT 27-FEB-1996 (first entry)
DE PML2 genomic DNA.
KW Human; thrombopoietin; TPO; mpl ligand; hml; fragment polypeptide;
  megakaryocytopoietic cytokine receptor; thrombopoietic signal;
  EPO domain fragment; erythropoietin; hEPO; haematopoietic cell;
  megakaryocyte; thrombocytopenia; myeloproliferative disease;
  inflammatory thrombocytosis; iron deficiency; Epo; platelet;
  red blood cell; progenitor; hml-2; ss.
OS Sus scrofa.
FH Key Location/Qualifiers
FT CDS 1..987
FT /tag- a
FT /product- PML2
FT GB2285446-A.
FT 12-JUL-1995.
FT 21-DEC-1994; 025831.
FT 03-JAN-1994; US-176553.
FT 21-JAN-1994; US-185607.
FT 15-FEB-1994; US-196689.
FT 04-APR-1994; US-223263.
FT 25-MAY-1994; US-249376.
FT 02-DEC-1994; US-348658.
FT 02-DEC-1994; US-348657.
PA (GETH ) GENENTECH INC.
PI Eaton DL, de Sauvage FJ;
  WPI: 95-234018/31.
DR P-PSDB; R76170.
PT Thrombopoietin polypeptide, ligand for mpl cytokine receptor -
  useful for treating thrombocytopenia and related diseases
  PS Example 13; Fig 21; 192pp; English.
CC This sequence represents a genomic clone encoding an isoform of porcine

```

CC thrombopoietin (TPO), also known as mpl ligand (PML), PML2. This
CC sequence was isolated by RACE PCR using the primer sequences given in
CC 094140-43 and 094128. The genomic sequence was isolated from a porcine
CC genomic library in EMBL3 by screening with pR45. Two forms of PML were
CC isolated in this manner. The first was a full length clone (see also
CC 094112) and the second encoded a protein with a four amino acid deletion
CC (PML2). Comparison of the sequences of PML and PML2 shows that the
CC latter form is identical except for the deletion of the tetrapeptide QLRP
CC corresponding to residues 111-114. The four amino acid deletions
CC observed in murine, human and porcine M1 cDNA occur at precisely the
CC same position in the predicted proteins (see also 094110). PML binds to
CC mpl which is a megakaryocytopoietic cytokine receptor which may transduce
CC a thrombopoietic signal.
SQ Sequence 1014 bp; 216 A; 365 C; 220 G; 213 T;

094113 Length: 1014 April 28, 1998 12:51 Type: N Check: 5829 ..

```
1  AGCCGGGCTC CTCCTGCTTG TGACCCCGCA CTCCTAATA AACTGCTTCG
51  TGACTCCCAT GTCTTCACG GCAGACTGAG CCAGTGCCA GACATTAAAC
101  CTTGTGCCAC ACCTGTCTTG CTGCCTGCTG TGGACTTAC CTTGGAGAA
151  TGGAAACC AGACGAGCA GACAAAGCA CAGATGTCC TGGAGGCCAC
201  AACCTTCTG CTGAGGACG TGATGACAG ACGGGACAA GTGGAGCCCC
251  CTTCCTCTC ATCCCTGCTG GTGCAGCTT CTGACAGAGT TCGCCTCTC
301  CTCGGGGCCC TGCAGGACCT CCTTGAATG CAGGGAAGA CCAAGCTCA
351  CAAGGATCCC AGTGCATCT TCTGAACCT CCAACAACG CTCGAGAGAA
401  AGGTGGCTTT CCTGCTCCTT GATGAGGGG CCTCCCTCTG TGCCAAGAG
451  GCCCCACCG CCATAGCTGT CCCGAGCAGC ACCTCTCCAT TCCACACACT
501  GAACAAAGTC CCNACAGGA CCTTGATTT GTGAGACA AACTCCAGTA
551  TCTCAGCCAG AACTACTGGC TCTGATTTT TCAAGAGCT GCAGGCATTTC
601  AGAGCCAGA TTCCTGCTT GCTGAACCA ACCTCCAGT CCTAGACCA
651  AATCCCTGGA CACAGAGATG GGACACAGG ACCCTGAGT GGAATTCATG
701  GACTCTTTC TGAACCCAA CCGGGGGCC TCGSAGCTC AGACATTCTT
751  CCAGCAACTT CAGGCAATGG CTCCGGGCA ACCTACCTCC AGCCTGAGA
801  GTCTCTTCC CCAGTCACG CTTCTCTTG AGATACACT CTCTCTCTC
851  CTTACCCAC CTGCCCCCTC CCCACAGTC AGCTCCAGC TCTGCTTCT
901  GACCCCTCTG CGATCACACC CAATCTACG AGTCTCTTC TATTGACAG
951  TCACCTTCAT TTCAGAGACC TGTCTCAGA AGAGTAAGT GCTCAGACC
1001 TGCCAACCTC AGCA
```

 WISE (TM)

Release 3.0.4A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1997 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MSearch_n n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Fri Apr 24 08:28:34 1998; MasPar time 501.50 Seconds

Tabular output not generated. 1407.620 Million cell updates/sec

Title: >US-08-483-534A-1
 Description: (1-597) from US08483534A.seq (2 of 2)

Perfect Score: 597
 N.A. Sequence: 1 TACCCCTGCGCCGTAAGAAAC.....TGAAGGGGGGACATTAAGC 597

Comp: ATGGGAGAGGGACCTTTTGTG.....ACTTCCCGCCCTTGATACG

Scoring table: TABLE default
 Gap 6

Mismatch STD : Dbase 0; Query 0

Searched: 354530 seqs, 591221332 bases x 2

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database:

emb13
 1:em_un 2:em_or 3:em_com 4:em_ov 5:em_pl 6:em_htg
 7:em_hum1 8:em_hum2 9:em_da 10:em_ro 11:em_un 12:em_v1
 13:em_pat
 genbank105

Database:
 14:gb_ro 15:gb_com 16:gb_ov 17:gb_in 18:gb_pl 19:gb_da
 20:gb_st 21:gb_v1 22:gb_ph 23:gb_sy 24:gb_un 25:gb_pat
 26:gb_htg

Statistics: Mean 10.301; Variance 5.091; scale 2.023

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
C 1	82	13.7	21948	17	CEP58B3	5.20e-41
C 2	82	13.7	85953	26	CEH04L24	5.20e-41
C 3	82	13.7	131318	26	CEV42E10	5.20e-41
C 4	81	13.6	498	23	ASEMAP1	3.09e-40
C 5	73	12.2	1068	14	MM010118	4.27e-34
C 6	73	12.2	1086	25	I50092	4.27e-34
C 7	72	12.1	7218	25	I66494	2.46e-33
C 8	66	11.1	1703	18	AF021800	8.33e-29
C 9	34	5.7	12325	18	SCV11EFT	2.51e-06
C 10	34	5.7	12325	18	SCV11EFT	2.51e-06
C 11	34	5.7	12325	18	SCV11EFT	2.51e-06
C 12	32	5.4	1745	18	SCU1348	4.42e-05
C 13	31	5.2	215	25	I28278	1.80e-04
C 14	30	5.0	11157	19	ECAE000389	7.22e-04

C 15	30	5.0	11000	19	ECOW67	7.22e-04
C 16	28	4.7	215	25	I28278	1.09e-02
C 17	26	4.4	1663	15	MVU92534	1.49e-01
C 18	26	4.4	10558	19	AE001003	1.49e-01
C 19	25	4.2	1131	17	DROHICP02	5.28e-01
C 20	25	4.2	3916	17	DROHGT4P13	5.28e-01
C 21	25	4.2	3991	17	DROHGT3P12	5.28e-01
C 22	25	4.2	4222	17	DROHGT2P11	5.28e-01
C 23	25	4.2	4297	17	DROHGT1PA	5.28e-01
C 24	24	4.0	450	16	CHCKM1	1.82e+00
C 25	24	4.0	4500	17	CEATOP2	1.82e+00
C 26	24	4.0	10772	17	AE012089	1.82e+00
C 27	23	3.9	565	25	E04076	6.08e+00
C 28	23	3.9	1000	14	NRNRNMOBP	6.08e+00
C 29	23	3.9	1951	14	MUSRP77B	6.08e+00
C 30	23	3.9	2405	14	RNMOBP81	6.08e+00
C 31	23	3.9	2486	14	RATMOBP2	6.08e+00
C 32	23	3.9	3442	14	RNMOBP81P	6.08e+00
C 33	22	3.7	69	25	I41362	1.96e+01
C 34	22	3.7	722	14	MUSCALMDA	1.96e+01
C 35	22	3.7	1288	19	SPEMM41	1.96e+01
C 36	22	3.7	3113	1	DMR81432	1.96e+01
C 37	22	3.7	3699	17	DMAC000667	1.96e+01
C 38	22	3.7	7869	17	DMKETE1	1.96e+01
C 39	22	3.7	25646	17	PVU51723	1.96e+01
C 40	22	3.7	41088	17	CEH0280	1.96e+01
C 41	22	3.7	82393	17	AC003923	1.96e+01
C 42	22	3.7	98334	26	AC003102	1.96e+01
C 43	22	3.7	128614	26	HS63G5	1.96e+01
C 44	21	3.5	24570	17	CELT05E8	6.10e+01
C 45	21	3.5	186388	26	AC003694	6.10e+01

ALIGNMENTS

RESULT	LOCUS	DEFINITION	ACCESSION	NID	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	COMMENT		
1	CEP58B3	Caenorhabditis elegans	21948 bp	DNA	INV	23-JAN-1998	Caenorhabditis elegans	1 (bases 1 to 21948)	Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berts, M., Bonfield, J., Burton, J., Connell, M., Cope, T., Cooper, J., Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A., Fulton, L., Gardner, A., Green, P., Hawkes, T., Hillier, L., Jier, M., Johnston, L., Jones, M., Kersey, J., Kirtlen, J., Laister, N., Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B., O'Callaghan, M., Parsons, J., Percy, C., Rifkin, L., Roop, A., Saunders, D., Showkeen, R., Smaildon, N., Smith, A., Sonhammer, E., Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M., Vaughan, K., Waterston, R., Watson, A., Weinstock, L., Wilkinson-Sproat, J., and Wohlman, P.	2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans	Journal Nature in press (bases 1 to 21948)	Direct Submission	Submitted (13-MAY-1996) Louis, MO 63110, USA. E-mail: jee@umsl.ac.uk or rvenematode.wustl.edu
2	CEH04L24	Caenorhabditis elegans	85953 bp	DNA	INV	23-JAN-1998	Caenorhabditis elegans	1 (bases 1 to 85953)	Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berts, M., Bonfield, J., Burton, J., Connell, M., Cope, T., Cooper, J., Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A., Fulton, L., Gardner, A., Green, P., Hawkes, T., Hillier, L., Jier, M., Johnston, L., Jones, M., Kersey, J., Kirtlen, J., Laister, N., Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B., O'Callaghan, M., Parsons, J., Percy, C., Rifkin, L., Roop, A., Saunders, D., Showkeen, R., Smaildon, N., Smith, A., Sonhammer, E., Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M., Vaughan, K., Waterston, R., Watson, A., Weinstock, L., Wilkinson-Sproat, J., and Wohlman, P.	2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans	Journal Nature in press (bases 1 to 21948)	Direct Submission	Submitted (13-MAY-1996) Louis, MO 63110, USA. E-mail: jee@umsl.ac.uk or rvenematode.wustl.edu

or longer because we arrange for a small overlap between neighbouring submissions.

The true left end of clone F5883 is at 1 in this sequence. The true left end of clone ZK809 is at 21836 in this sequence. Coding the sequences below are predicted from computer analysis, using the program Genefinder (P. Green, ms in preparation), and other available information.

The end of this sequence (21836..21948) overlaps with the start of sequence CEZK809.

FEATURES

```

source
    location/Qualifiers
    1..21948
    /organism="Caenorhabditis elegans"
    /db_xref="taxon:6239"
    /chromosome="IV"
    /clone="F5883"
    complement(join(553..914,967..1094,1525..1892,1939..1989))
    /codon_start=1
    /product="F58B3.8"
    /db_xref="PID:e244447"
    /db_xref="PID:g1370035"
    /translation="MQLATSRFLPFIPIISICGYLEIHIOSPHIOATLVNISNF
    PRKSKIVHKLKSNPENTRYLPVNFNDINITIKGKSEDEGKRSYKNTPKYL
    GLDITIVTHPRVGLSPAAVARPLNGIRIVIGCKNMFGEKCEFCEDNATNSG
    KHCNFGKPSCLNGITGSKCEKLIIVPLCEGVCIDSNPRPICESCAYGEKCE
    KKHISASLSHYKGFEMFEMTFLFMSLVENELAEIKSEAWETINMLGNVSDT
    PFANRPNTETPTGCVNLNDVDII"
    complement(join(2547..2989,3948..4149))
    /note="protein predicted using Genefinder; Weak similarity
    to E.histolytica lysozyme (PIR Acc. No. S55506)"
    /codon_start=1
    /product="F58B3.1"
    /db_xref="PID:e244554"
    /db_xref="PID:g1370028"
    /translation="MKRFVSALLIGASAARNGIDFIOVPVSTATFCQIHKAGSYVI
    PRFPTSVGLTHDGIQNVKARAAGLVDVGIYFPCIRSCPSANQYKTDALMSA
    GHVSITMLDIERLAMPANASNRATLEAVRQAQVAKQGVGITSNTNMQDITVGLDY
    HGSSITMLWATYDGIKDFSKAPFGGWSKPTIHQWEGITTGCGVSDMNYVP"
    complement(join(5347..5789,6421..6625))
    /note="Weak similarity to E.histolytica lysozyme (PIR Acc.
    No. S55506)"
    /codon_start=1
    /product="F58B3.2"
    /db_xref="PID:e244555"
    /db_xref="PID:g1370031"
    /translation="MKHFFITLFCGSVSAARNGIDINSVSTSTFTCIKNAGSFI
    IPIHYSSGVDTGVQNVKARAAGLTDVGYIFPCIKSTCASANQYKASIDAKKA
    AGKISTLMIDIERLAMPADHASNRATLEAVSEAKAYGQGVGITSNTNMQDITVGLD
    YHGQSSCLMWPAIDGVKDFSKYAPFGGWSKPTIHQWSDTTTGCGVSDMNYIP"
    complement(join(8013..8455,8785..8986))
    /note="Weak similarity to E.histolytica lysozyme (PIR Acc.
    No. S55506); cDNA EST YK94d3.3 comes from this gene; cDNA
    EST YK94d3.5 comes from this gene"
    /codon_start=1
    /product="F58B3.3"
    /db_xref="PID:e244445"
    /db_xref="PID:g1370032"
    /translation="MKLVSTLILISAVSARNGIDFIOVPVATFTCVKNAGSYVI
    PRFPTSVGLTHDGIQNVKARAAGLTDVGYIFPCIKSTCASANQYKASIDAKKA
    GRTISTLMIDIERLAMPADHASNRATLEAVSEAKAYGQGVGITSNTNMQDITVGLDY
    HGQSSITMLWAAIDGVKDFSKYAPFGGWSPTIHQWEGITTAGCGVSDMNYVP"
    complement(join(9873..10664,10717..10897,10953..11093,
    11447..11539,11655..11759,11810..12324))
    /codon_start=1
    /product="F58B3.4"
    /db_xref="PID:e244556"
    /db_xref="PID:g1370033"
    /translation="MTLHLETFNFQVVIDKRFPAALTDREFTRAKVDMRGRKQKT
    VGNAMLIDIEDEEYQPLKKEIKSKTTKDDDELDFEDDDEDEADDEDEDEDEDE
    DEISDEENDEEENKIGMNGFRKLARGGVNDSDDDSDEDEGDEVEDDEDE
    GGIEDLANIDREVDQVETSNLAVCNLEMEDNCEIDILIKSYENRKKTFKNNL
    KKRENTFIETVDFERIEKPOGGSIVSGIYSDGKGLSKSEETGTPLLAKTAVE
    EKREDEMDDETATVREYLVNKLKYYSVTTPGISAVAVYECGQFQFEELGLMD
    MRFIPDMDFEEDRVEKFLNAEDVNLTKYAKKKKSAISTGAKILMDEDDPQRKKK

```

CDS

```

FLEAFNGDADGKGLIVDSDESDGEANRKTIALMLNTDERSKLDVDMEGEKNGSD
SSDGEYVAYDDDEEIGVKKKKKGSDDEDEEVEEKEEDVLTGYAYKKQKOK
LMENRKLKRGKTSKEATNTIKTYAADSISKDRDFSLPDSYALIEPSSKRTGSLY
TKOAEKSKSGSSVITETKPEDVQKQADQADNRKKSVKN
complement(join(13218..13349,13423..14081,14133..14237,
14568..16021,16068..16391,16444..16523))
/note="Similarity to yeast methionyl-tRNA synthetase
(SW:SYMC_YEAST): cDNA EST YK9d5.5 comes from this gene;
cDNA EST YK11b4.5 comes from this gene; cDNA EST YK9d5.3
comes from this gene; cDNA EST YK11d4.3 comes from this
gene; cDNA EST YK112d5.5 comes from this gene; cDNA EST
YK190e7.3 comes from this gene; cDNA EST YK190e7.5 comes
from this gene; cDNA EST YK135d9.3 comes from this gene;
cDNA EST YK135d9.5 comes from this gene; cDNA EST
YK472h4.3 comes from this gene; cDNA EST YK472h4.5 comes
from this gene; cDNA EST YK461h12.3 comes from this gene;
cDNA EST YK461h12.5 comes from this gene; cDNA EST
YK354c6.3 comes from this gene; cDNA EST YK354c6.5 comes
from this gene; cDNA EST YK314e7.3 comes from this gene;
cDNA EST YK314e7.5 comes from this gene; cDNA EST
YK287d5.3 comes from this gene; cDNA EST YK287d5.5 comes
from this gene"
/codon_start=1
/product="F58B3.5"
/db_xref="PID:e244446"
/db_xref="PID:g1370034"
/translation="MGHDADITKKSFEASLPGYVEKKDKPSILIPQGRNILLTALP
YNNVPHLNTIGCVLSADVFATYCNLRHQHFFYGGIDEGIGATETKALQEGCPRE
LCDCKHAIRKGIYEMFGIDFSHGRTDDHGGELQDMALDKHKNYTSQSVOLYC
NOCERKLADRPVYGTGCMCAVDDARDDQDGGKILNADVLDAKCHMKAKPEVOS
THIFLSDLRLOOQTEHDLRELAKEENRNSNAVQKAMKGLDPRCITRDLKMG
AYLDGEKKEKPYVWEDAPIGYLSITKCYLGNNMTKWNKEVLEFNFVGDNVAF
WAMFPCSOIGANDNYVNNVLCATELYANEDPKFSKRSRTGIFGDAOCTELPADIR
PULIYRPSQOTAFSMDPFYKANSSELNLCNTINRLNLSVANSFGCVYEDMNTIR
DDEAVLSETHNECOMDKOPDVHLDAKATILNLSRLNQMQATVPVNLKDEBG
KRAAGITIGVANAIVHVSLLYPIPIPTISATIRECGPAPLFLPPICYLKGKHK
ICQPSLEIPFLQPAQIAEFKARFGGSDSSAPKTAERKQKQKAPRDKGDKM
ASTAAVELEOGAKVLSOLIAONLKFKDOAKLFTENOQRIDGEKQGLITVKTJQH
OLIELETTAGTIOGKVPYVSCPTPTSPASGITTEAPKKAESPPEPKAKQK
KKGKGAAPVPDITDVGDLDMRVGIIICEKHDPDADALYQDIDVGEASRPTVYSGL
VRAVPLDQNRLLVVVLCNLPKPKKNGVSRAMVWCASSPDVVEIYNEVADSRTGTV
VCPPTTHRPDEQLNPKKIMETVAEDLKVSAAGFAEMKQPLLIGSESEMTAPTLGV
HYK"
join(17716..17921,17974..18388,18443..18919,18977..19057)
/note="protein predicted using Genefinder; Similarity to
mouse interferon-related protein PC4 (SW:PC4_MOUSE); cDNA
EST cm01c9 comes from this gene"
/codon_start=1
/product="F58B3.6"
/db_xref="PID:e244443"
/db_xref="PID:g1370029"
/translation="MGRNNRKKKDDPYIGPAKTGRNDSDDSSASATYMEDIOSI
MGDVEDMNDVYDQVLDNRADQDKNSTYIDGLHRYNLLVTKPAEPETNRKKTITMA
VCSRAMKPDPSAOFIARLVGLVAVOAGEISEMLAEPMSQKRTILMDSRCVFLTA
CATSLAIVTRICASEDEEYVANAACAKRFAHATKSGSANDGHATLINDVADDPK
ITVILSSNOLLEVRLAAETLAFLEHMOARAGFPKNEHVLSDLDQWVNSSKRT
KDKDKRQRAVADPIRAFITADEDDAPVNNKIQOQTSLDSCAIKIFVDTCCLLHGL
AQQLLHNEVLRVDFDLGALPBAESANKQSLRAVHEADKRRNVRGQRKRSTVY"
join(19370..19424,19481..19563,20023..20835,20888..21052)
/note="protein predicted using Genefinder; Similarity to
Arabidopsis DNA-damage-repair protein DRL11
(SW:D11_ARATH)"
/codon_start=1
/product="F58B3.7"
/db_xref="PID:e244444"
/db_xref="PID:g1370030"
/translation="MYNDEDDYPOGPPAKQAKPWNHTOMAFMOSQALORAKAQQO
AAROKLYSSAPPVYIDLSRRTITTTATSSQFORIPRANPSDINSIFPKATDESV
MAGEBHVCETIPYTPNTEYVAAREINRQKRETKARVARLRLDEEEDKRSKG
AAIAPPTMLMEPEDEPVTKNTNDEKPPSPSPSPFAPAGKATSRGLGAAIIMK
RGYVEGALGISEGMSKSTALIERGTGAGGIVAEAPAPAFATNSKMAVONAKYIL
QAMNLIDLSVEVSGEGRKEFADEIKEMEKCCGVVYVYHVDSEGEDRQVAVFEET

```



```

CP      344  GAAACACACAGAACACATCTTGGACACGACTCTCTCATTCTTGGGTTGAGGTTG 285
Db      44246 CAAGACACAGACAGAGCGGTTCTGCATTGTAGAGTGAAGTGAAGCAGTCCA 44305
CP      284  CACAGCACACATACACAGCTCTCTCGACATTCCTCTTGGGACAGCACTGACAGCGCCG 225
Db      44306 GAAACTACAGTTCGGGGGCTGACTCTCCAAACATGATCTGTCCAGCTAAAGACATCT 44365
CP      224  CTCACACAGTCCGTTGCTTACGCTTCCCCACAGTCATCTCTACATACAGGCTGTCT 165
Db      44366 GCATCAGAGATCTCTCTCAGCTTGTATGTGACCAAGCGCATATCCAG 44415
CP      164  GCATCTGGGCTCTCTCCACAGATGATTTTCCACACAGCATATCCAG 115

RESULT  4
LOCUS   AEMAPII 498 bp DNA SYN 10-OCT-1997
DEFINITION Artificial sequence DNA for cytokine EMAPII.
ACCESSION Y08737
NID      g1619591
KEYWORDS EMAPII gene.
SOURCE   synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 498)
AUTHORS Schluesener H.J., Seid, K., Zhao, Y. and Meyermann, R.
TITLE    Localization of endothelial-monocyte-activating polypeptide II
          (EMAP II), a novel proinflammatory cytokine, to lesions of
          experimental autoimmune encephalomyelitis, neuritis and uveitis:
          expression by monocytes and activated microglial cells
          Glia 20 (4), 365-372 (1997)
          97404153
          2 (bases 1 to 498)
          Schluesener H.J.
          Direct Submission
          Submitted (01-OCT-1996) H.J. Schluesener, Institute of Brain
          Research, Calwer Str. 3, D-72076 Tuebingen, FRG
FEATURES
    source
        1. 498
        /organism="synthetic construct"
        /db_xref="taxon:32630"
        1. 498
        /gene="human EMAPII"
        <1..>498
        /gene="human EMAPII"
        /codon_start=1
        /db_xref="PID:e274145"
        /db_xref="PID:g1619592"
        /translation="SKPIDVSRDLRIGCIITARKHPDADSLVEEVDGEIAPRTV
          SGIVNHYPLEOMONBMVTLCLNKPARKRGVLSQAMWCASSPEKIIILAPNGSVPG
          DRITFAFPSEPKELNPKKIMQIOPDLHTNDECAVATKGVFEVKGAGVCAQTM
          SNGGIR"
    gene
        /db_xref="taxon:32630"
        1. 498
        /gene="human EMAPII"
        <1..>498
        /gene="human EMAPII"
        /codon_start=1
        /db_xref="PID:e274145"
        /db_xref="PID:g1619592"
        /translation="SKPIDVSRDLRIGCIITARKHPDADSLVEEVDGEIAPRTV
          SGIVNHYPLEOMONBMVTLCLNKPARKRGVLSQAMWCASSPEKIIILAPNGSVPG
          DRITFAFPSEPKELNPKKIMQIOPDLHTNDECAVATKGVFEVKGAGVCAQTM
          SNGGIR"
    CDS
        1. 498
        /organism="synthetic construct"
        /db_xref="taxon:32630"
        1. 498
        /gene="human EMAPII"
        <1..>498
        /gene="human EMAPII"
        /codon_start=1
        /db_xref="PID:e274145"
        /db_xref="PID:g1619592"
        /translation="SKPIDVSRDLRIGCIITARKHPDADSLVEEVDGEIAPRTV
          SGIVNHYPLEOMONBMVTLCLNKPARKRGVLSQAMWCASSPEKIIILAPNGSVPG
          DRITFAFPSEPKELNPKKIMQIOPDLHTNDECAVATKGVFEVKGAGVCAQTM
          SNGGIR"
BASE COUNT 130 a 111 c 147 g 109 t 1 others
ORIGIN
Query Match 13.6%; Score 81; DB 23; Length 498;
Best Local Similarity 67.5%; Pred. No. 3.09e-40;
Matches 156; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

```

```

Db      201  GAAACCGCGGAAATGCGTGGCTGCTGTCTCAGGCAATGCTGATGCGC 251
Oy      291  GAAACCCAGAAATGAGAGAGATCGAGTCCCAAGGATGCTGTGATGTC 341

RESULT  5
LOCUS   MMU10118 1068 bp mRNA ROD 14-FEB-1995
DEFINITION Mus musculus endothelial-monocyte activating polypeptide II mRNA,
ACCESSION U0118
NID      g4498911
KEYWORDS house mouse.
SOURCE   Mus musculus.
ORGANISM Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
          Vertebrata; Euthera; Rodentia; Sciurognathi; Myomorpha; Muridae;
          Murinae; Mus.
REFERENCE 1 (bases 1 to 1068)
AUTHORS Kao, J., Houck, K., Fan, Y., Haehnel, I., Libutti, S. K., Kayton, M. L.,
          Grikscheit, T., Chabot, J., Nowyrod, R., Greenberg, S., Kuang, W.-J.,
          Leung, D. W., Hayward, J. R., Kiesel, W., Heath, M., Brett, J. and
          Stern, D. M.
TITLE    Characterization of a novel tumor-derived cytokine.
          Endothelial-monocyte activating polypeptide II
          J. Biol. Chem. 269 (40), 25106-25119 (1994)
          95014290
          2 (bases 1 to 1068)
          Kao, J., Fan, Y., Haehnel, I., Brett, J., Greenberg, S., Clauss, M.,
          Kayton, M., Houck, K., Kiesel, W., Seifeld, R., Burnier, J. and
          Stern, D.
          A peptide derived from the amino terminus of
          endothelial-monocyte-activating polypeptide II modulates
          mononuclear and polymorphonuclear leukocyte functions, defines an
          apparently novel cellular interaction site, and induces an acute
          inflammatory response
          J. Biol. Chem. 269 (13), 9774-9782 (1994)
          94193665
          3 (bases 1 to 1068)
          Kao, J., Ryan, J., Brett, G., Chen, J., Shen, H., Fan, Y., Godman, G.,
          Familletti, P. C., Wang, F., Pan, Y. E., Stern, D. and Clauss, M.
          Endothelial monocyte-activating polypeptide II. A novel
          tumor-derived polypeptide that activates host-response mechanisms
          J. Biol. Chem. 267 (28), 20239-20247 (1992)
          93015897
          4 (bases 1 to 1068)
          Houck, K. A.
          Direct Submission
          Submitted (27-MAY-1994) Keith A. Houck, Molecular Biology, Sphinx
          Pharmaceuticals Corp., P.O. Box 52330, Durham, NC 27717, USA
FEATURES
    source
        1. 1068
        /organism="Mus musculus"
        /db_xref="taxon:10090"
        /cell_line="Weth A"
        /cell_type="fibrosarcoma"
        1. 1068
        /standard_name="EMAP II"
        /codon_start=1
        /product="endothelial-monocyte activating polypeptide II"
        /db_xref="PID:g498912"
        /translation="MATNDVAVLRLEKROKGAENDQIEYLKQVALLKEKRIIQAATRE
          EKLREVMKRLKEIELEKELIILAHNGVEVRVRLSTNCTGASVVSQSPV
          ATASPRATNQIRAGEEKYKTEKEKEKEKQQAASSTDSKPIDASRLDRIGCI
          VTKKHPDADSLVEEVDGEAPRTVSELVNHVPLEOMONBMVTLCLNKPARKRG
          VLSQAMWCASSPEKVEIILAPNGSVPGDRITFDAPGEPDEKLNPKKIMQIOPDL
          HTNAECVATYKGAPEVKGAGVCAQTMANSIGIK"
          496..993
          /standard_name="EMAP II"
          /product="endothelial-monocyte activating polypeptide II"
BASE COUNT 328 a 212 c 297 g 231 t
ORIGIN
polyA_signal
1037..1042
328 a 212 c 297 g 231 t

```


	/db_xref="PID:g2754814"
	/translation="MCEYFRRLPGTAPGSSPPAPATHRLILGRELRYVRMFICREF AKAATNDVLRLEOKGEADQIILEYLKQVALIKERAVLQATLREKKRLEVNAKLKE KIELLOQLQIAETIQNGVKQIPVPQSDITPVQASSAVSSTVISOSTSVTISCISKEP SKGGEEKRKYAEITDKGCKEKKELQSPADSPVDSPLDLRIGITIVVKKHPD DSLVEEDVGEAAPRTYISGLNVHNPIDDMONRVYLCKQPKAKMGCTLSOAMWKC ASSPEEIIAPNGSVSGDRITFDFAFGEDEKELNPKRKIMWEILOPLHNACVAT YKSPDFEVKGKGVCAQTGMANSGLK"
BASE COUNT	400 a 251 c 343 g 260 t
ORIGIN	
Query Match	11.1%; Score 66; DB 14; Length 1254;
Best Local Similarity	64.1%; Pred. No. 8 32e-29;
Matches	150; Conservative 0; Mismatches 84; Indels 0; Gaps 0;
Db	681 TCTCGTCGTGATCTTCGAATTGGTGTAATTGTTACTGTCAAAAAGCACCCGAGCAGAC 740
Qy	109 TCCGGCGTGATATCCGTGTGGGAAATCATCACTGTGGAGAACAACCACAGATGCAGAC 168
Db	741 TCACGTGTGTAGAAGAAAGTAGTGTGGGGAGGCCAACCCCCAAGACAGTATCATAGTGA 800
Qy	169 AGCCGTATGTAGAAATAATTGACGTGGGGGAAGCTGGAACCGGACACTGTGTGAGCGGC 228
Db	801 CTGGTGAATCATGTCTCTAGACCAAGATGCAAAATCGAATGGTGTTTTACTTGTAT 860
Qy	229 CTGTGACAGTTCGTGTCGCCAAGAGSAGAACTGCAGACAGAGCTGTGTGTGTGTGCAAC 288
Db	861 CTGAAGCCTGCAAAAAATCGGGGAATACTGTCTCAAGCAATGGTGATGTGTCT 914
Qy	289 CTGAACCCCAAGAGATGAGAGAGTGCAGTCCCAAGCAGCATGCTCTGTGTGTCT 342
RESULT	9
LOCUS	SCYGL105W 1703 bp DNA PLN 11-AUG-1997
DEFINITION	S.cerevisiae chromosome VII reading frame ORF YGL105W.
ACCESSION	Z72627 Y13135
NID	G1322647
KEYWORDS	baker's yeast.
SOURCE	Saccharomyces cerevisiae
ORGANISM	Eukaryote; mitochondrial eukaryotes; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces. 1 (bases 1 to 1703) Castagnoli,L., Paoluzi,S. and Manenkova,O. Unpublished 2 (bases 1 to 1703) MIPS. Direct Submission Submitted (14-MAY-1996) Data collected by MIPS on behalf of the European yeast chromosome VII sequencing project. MIPS at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152 Martinsried, FRG; E-mail: Mewes@mips.emblnet.org Location/Qualifiers 1. .1703 /organism="Saccharomyces cerevisiae" /db_xref="taxon:4932" /chromosome="VII" 428. .1558 /gene="GAP1" 428. .1558 /gene="GAP1" /note="ORF YGL105W" /codon_start=1 /db_xref="PID:e243983" /db_xref="PID:g1322648" /translation="MSDLVTKEFSLLISKYPSTFEKSQAQAQWESVLKSGIOPHL DOALNLVRNTFLVSLTPSTDVHFVEVALPIKIDLVASSKVKSTYTRHLMI DYMNOLLEVSSTDKLEINHDLDPHEVIERRKAAPGAGADAARADEVDYSKKAKOD HPHGCDPEFTLKREEAKRAKKAANAANKOOBONKAREKPSAIDRFVGFQK AIKHPPADSLYSTIDVGDEEGPRYCSGVKRPFPDAQGERVVVVVCINKPYNMGI KSIAMVLCGSNDKVFEFEPKDSKAGDKVFEEFGFDEVAEMQNLRKKIWEILOPH
FEATURES	
SOURCE	
gene	
CDS	

BASE COUNT	577 a	318 c	335 g	473 t	TTNDGLEVFIFKDEEKDHPVVRKLTINAKSESEFKVASIANNQV"
ORIGIN					
Query Match			5.78;	Score 34;	DB 18; Length 1703;
Best Local Similarity	59.58;	Pred. No. 2.51e-06;			
Matches 154; Conservative	0;	Mismatches 102;	Indels 3;	Caps 1;	
Db	1028	CCAGAAAGCCCAAGACCCCAATAGATTTCCGTTGGTTTATTTCAAAAGCAGC	1087		
Qy	91	CCAGAGAGAGTATCCATCCCGCGCTGGATATCCGTGGGCAAAATCATCTGTGGAG	150		
Db	1088	AAGCACCAGATGCGCAGCTCCTATATGTCTCTACATATGATCGGTGATGAAGAGC	1147		
Qy	151	AAGCACCAGATGACAGACAGCTGTATGAGAAGATGAGGTGGGGAAGCTGAA---	207		
Db	1148	CCAAACTGTTGTTGTGTGGTTGGTCAAGCATTTTCCCTTTGGACCTTATGCAAGACGT	1207		
Qy	208	CCAGGACTGTGTGTAGGCGGCTGTACAGTTCGTGCCCCAAGAGAGACGTGCAGGACAG	267		
Db	1208	TATGTGTTGTGTATGTAACACTTGAACAGTATACATGAGAGGTATTAATCACTGCT	1267		
Qy	268	CTGTGATGTGTCTCTGTCAACTGAAACCCAGAAAGATGAGAGCTGATGTCACAGGC	327		
Db	1268	ATGATGTTGTGTGTGTTCTTA 1286			
Qy	328	ATGCTTCTGTGTGTTCTTA 346			
RESULT	10				
LOCUS	SCAR1	2473 bp	DNA		14-OCT-1996
DEFINITION	S.cerevisiae ARc1 gene.				
ACCESSION	X95481				
NID	g1620459				
KEYWORDS	ARc1 gene.				
SOURCE	unidentified.				
ORGANISM	unclassified.				
REFERENCE	1 (bases 1 to 2473)				
AUTHORS	Simos,G., Segref,A., Fasolo,F., Hellmuth,K., Shervenko,A.,				
TITLE	Mann,M. and Hurt,E.C.				
JOURNAL	The yeast protein Arc1p binds to tRNA and functions as a cofactor				
MEDLINE	EMBO J. 15 (19), 5437-5448 (1996)				
REFERENCE	97050848				
AUTHORS	2 (bases 1 to 2473)				
TITLE	Simos,G.				
JOURNAL	Direct Submission				
COMMENT	Submitted (01-FEB-1996) G. Simos, University of Heidelberg,				
FEATURES	Institut fuer Biochemie I, Im Neuenheimer Feld 328, 69120 Heidelberg, FRG				
SOURCE	Related sequence: U31348.				
location/Qualifiers					
1..2473	/organism="unidentified"				
/db_xref="taxon:32644"					
1..2473	/organism="Saccharomyces"				
/strain="yU4.2XJR26.198"					
/db_xref="taxon:4930"					
534..1684	/gene="ARC1"				
554..1684	/gene="ARC1"				
/codon_start=1					
/product="Arc1p"					
/db_xref="PID:g1620460"					
/translation="MSDIYMKPESLIISKPYSTFKQSAQAOMESYVLSGQIQPHI					
DQNLVLRDNTFIVSTILPTSTIDVHFVEVALPLIKDLVASSDKVSTYTYRHLRLWID					
DIYQNLLEVSSTDKLEINHDLPHVEYTEKKRPAAGADAADAAKADSDVSKAKKODD					
HPGKDEFTLKILREBAKAKAKAANKAANKQOENKAPKPKPAIDPRVGFQIR					
AIHNPADSLIVSTIDVDEEGPRTVCSGLVKEHPIDAMOEERYVYVCNKKVPMRGIT					

BASE COUNT	837 a	457 c	494 g	685 t
ORIGIN	KSTAMVLCGSDNDKVEEVEPEPRKSDAKDKNVFEESGDEAPRKQINPKKKIWEHQPHF TTNDGLEVIRFDEEEKDHPVKRLTNAGSESFVASIANAOVR"			

[illegible]

DB	10814	TATGTGTGTTGATGCAACTGAACAGTAACTAGAGAGGATTAATCACTGCT	10873
OY	268	CTGGAGTAGTGCTCTGCAACTGAAACCCAGAAAGATGAGAGTGCATGCCAAGGC	327
DB	10874	ATGCTATGTGTGTTCTTA	10892
OY	328	ATGCTTCTGTGTGCTTCA	346
RESULT	12		
LOCUS	SCU31348	1745 bp	DNA
DEFINITION	Saccharomyces cerevisiae G4p1 (g4p1) gene, complete cds.		PLN 02-OCT-1995
ACCESSION	U31348		
NTD	g1002711		
KEYWORDS			
SOURCE	baker's yeast.		
ORGANISM	Saccharomyces cerevisiae		
REFERENCE	Eukaryotes; mitochondrial eukaryotes; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.		
AUTHORS	1 (bases 1 to 1745)		
TITLE	Frantz,J.D. and Gilbert,W.		
JOURNAL	A novel yeast gene product, G4p1, with a specific affinity for quadruplex nucleic acids		
MEDLINE	J Biol. Chem. 270 (35), 20692-20697 (1995)		
REFERENCE	95386521		
AUTHORS	2 (bases 1 to 1745)		
TITLE	Frantz,J.D.		
JOURNAL	Direct Submision		
FEATURES	Submitted (10-JUL-1995) J Daniel Frantz, Molecular and Cellular Biology, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138, USA		
source	Location/Qualifiers		
gene	1..1745		
CDS	/organism="Saccharomyces cerevisiae"		
	/strain="5288c"		
	/db_xref="taxon:4932"		
	257..1384		
	/gene="g4p1"		
	257..1384		
	/gene="g4p1"		
	/note="G4 nucleic acid binding protein"		
	/codon_start=1		
	/product="G4p1"		
	/db_xref="PID:g1002712"		
	/translation="MSDLVTFEESLIISKYVSEFKESQAQAAQMEVLSKSGOTIPLDQLVLRNMFIVSTLYPTSDVHFEVALPLIKDLVASSKDYKSTYTRHLIRWIDYMONILEVSESTDLKLEINHDLDPHEVIERKKRLRVALLMOGSEDSKARKODPRGRPDEETLKLKEEAKALLAAKNAKKOQOQNKAPKRPASLDFVYGTQRIKIHADLASLYSTIDVDEDESPVCSGLVNHFPDVAQERYVVVVCCLKRVNMGISTAHADCSNDKVEFEPKPSRAGCSVFEEDGDEAPMKQLPKKIKIWEHLDPHFETNDGLEVIFKDEEKDHPVRKLTNAKESFEVSIANAQVR"		
BASE COUNT	582 a 326 c 345 g 492 t		
ORIGIN			
Query Match	5.4%;	Score 32;	DB 18;
Best Local Similarity	61.8%;	Pred. No. 4.42e-05;	
Matches 123;	Conservative	0;	Mismatches 73;
		Indels 3;	Gaps 1;
DB	914	AAGCACCAGATGCCGATCCCTATATGCTCTACAAATTGATGTCGGTGATGAAGAAGC	973
OY	151	AAGCACCAGATGCGAAGCCCTGATATGATAGAGATGTGAGCTGGGGAACCTGAA---	207
DB	974	CCAAAGACTGTTTGTGTTGTTGGTTCAGACATTTTCCTTGGAGCTATGCAAGACGT	1033
OY	208	CCAGCGATCTGCTGCTGAGGGCCGTGTACAGTTCTGCGCCCAAGAGAACTGCAAGCAG	267
DB	1034	TATGTTGTTGTTGATGCAACTGGAACAGTAAACAGTAAACATGAGAGATTAATCACTGCT	1093
OY	268	CTGGAGTAGTGCTCTGCAACTGAAACCCAGAAAGATGAGAGAGTGCATGCCAAGGC	327
DB	1094	ATGCTATGTGTGTTCTTA	1112

Oy 328 ATGCTTCGTGTGCTTCA 346
 RESULT 13 128278 215 bp DNA PAT 30-OCT-1996
 LOCUS DEFINITION Sequence 5 from patent US 5569830.
 ACCESSION 128278
 MID g1819054
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE
 1 (bases 1 to 215)
 Bennet,A., Labavitch,J.M., Powell,A. and Stotz,H.
 Plant inhibitors of fungal polygalacturonases and their use to
 control fungal disease
 JOURNAL Patent: US 5569830-A 5 29-OCT-1996;
 FEATURES Location/Qualifiers
 source 1..215
 BASE COUNT 15 a 8 c 25 g 26 t 141 others
 ORIGIN
 Query Match 5.2%; Score 31; DB 25; Length 215;
 Best Local Similarity 14.3%; Pred. No. 1.80e-04;
 Matches 22; Conservative 64; Mismatches 67; Indels 1; Gaps 1;
 Db 1 MTNVMSSSVSRASCDKAKKNDGNTSSWTDDCNRTNGVCDTDTYRRNNDGHNK 60
 Oy 395 GTGCTCCGTGTAGACACGTTGTGGAAGGCTATGAAGAGGCCAACCCAGATGAGGAC 454
 Db 61 YSSANYTGNNGVNAKAKTYHTHTVSGADSKVTDSYNASSTSSNGTDCNRSQADSY 120
 Oy 455 TCAACCCCAAGAAAGAGCTTCGAGAGCTTCAGAGCTCAAAATTTCTGAGAGAT 514
 Db 121 GSSKTAMTSRNFTGKTANNADVSRNMGASVGS 154
 Oy 515 GCATCGCACG-GTGGAGCAACCAACTTCATGA 547
 RESULT 14
 LOCUS ECAE000389 11157 bp DNA BCT 02-SEP-1997
 DEFINITION Escherichia coli K-12 MG1655 section 279 of 400 of the complete
 genome.
 ACCESSION AE000389 U00096
 MID g1789451
 KEYWORDS
 SOURCE Escherichia coli.
 ORGANISM Escherichia coli.
 REFERENCE
 1 (bases 1 to 11157)
 Blattner,F.R., Plunkett III,G., Bloch,C.A., Perna,N.T., Burland,Y.,
 Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
 Gregor,J.J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
 Mau,B. and Shao,Y.
 The complete genome sequence of Escherichia coli K-12
 Science 277 (5331), 1453-1474 (1997)
 JOURNAL MEDLINE 9742617
 REFERENCE
 2 (bases 1 to 11157)
 Blattner,F.R.
 TITLE Direct Submission
 JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
 608-263-7459
 3 (bases 1 to 11157)
 Blattner,F.R.
 TITLE Direct Submission
 JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
 608-263-7459

COMMENT

The E. coli K-12 sequence and its annotations have been updated.
 All of the ambiguous residues in our original submission have been
 resolved, and mis-assemblies in two repetitive regions have been
 realigned. The annotations have been improved and updated as well.
 With this release we begin designating a version number for the
 annotated sequence, to assist in keeping track of corrections,
 updates, and other changes. This is version M52 (SEPT. 02, 1997).
 In addition, a revised notation has been instituted which assigns
 each gene (protein- or RNA-encoding) a unique numeric identifier
 beginning with a lowercase 'b' (in the 'label' field); this will
 remain constant through further updates, gene identifications, etc.
 This sequence was determined by the E. coli genome project at the
 University of Wisconsin-Madison (Frederick R. Blattner, director).
 Supported by NIH grants HG00301 and HG01428 (from the Human Genome
 Project and NCHRR). The entire sequence was independently
 determined from E. coli K-12 strain MG1655.
 Predicted open reading frames were determined using Genemark
 software, kindly supplied by Mark Borodovsky, Georgia Institute of
 Technology, Atlanta, GA, 30332.
 e-mail: mark@amber.gatech.edu
 Open reading frames that have been correlated with genetic loci are
 being annotated with CG site Nos., unique ID nos. for the genes in
 the E. coli Genetic Stock Center (CGSC) database at Yale
 University, kindly supplied by Mary Berlyn. A public version of the
 database is accessible (<http://cgsc.biology.yale.edu>).
 Annotation of the genome is an ongoing task whose goal is to make
 the genome sequence more useful by correlating it with other data.
 Comments to the authors are appreciated. Updated information will
 be available at the E. coli Genome Project's World Wide Web site
 (<http://www.genetics.wisc.edu>).

FEATURES

source

1..11157
 Location/Qualifiers

CDS

/organism="Escherichia coli"
 /strain="K-12"
 /sub_strain="MG1655"
 /db_xref="taxon:562"
 174..797
 /note="0207: This 207 aa ORF is 42 pct identical (7 gaps)
 to 52 residues of an approx. 200 aa protein SLTD_ECOLI SW:
 P30856"

gene

/codon_start=1
 /label="b3071"
 /db_xref="pid:g1789452"
 /transl_table=11
 /translation="MSHHBCCCHBEGQPRHEGCKEKEKSHHEGCHGHQHEHQCCG
 GRHGRGGRORFEGHGLRVYIIDLISRDSSHVELIKAIENLTGQNPSPVITP
 TLDFLOEQSLTIREEGRKQIALTEGQAGMLEENREOVMIEERIKARCVGAALRO
 NPOKRALDNFKAVLDLRVNSDISDAQIKIIVIDRAAFDITQLD"
 complement(951..2471)
 /gene="air"
 complement(951..2471)
 /gene="air"
 /note="f506: 100 pct identical to AIR_ECOLI SW: P50466"
 /codon_start=1
 /label="b3072"
 /product="aerotaxis receptor protein"
 /db_xref="pid:g1789453"
 /transl_table=11

gene

/translation="MSSHPYVTOONTPLADDTLTMTDLSQSYTTHANDTFVQSGYT
 LOELQOGPHNNVRADPMKAFADMPYTLKGEPSGIVKRRKNGHGVYRAVAPM
 VREGKISGYSIFRTADEIEAIEPYTKALNAGRTSKRIHKGIVRGWGRKPSLP
 LMRARVMTLMTFTLLAAMLMFVAAPVYVTLAVYLLASACEPMQIVPIEYVAO
 ALKAVTGERNSVEHLNSDELGLTRAVGOLGLCMRWLINDVSCVSSVANGSTELAK
 GTDELNEHTOOTVDNVQVATNMQASVKONSATASADKLSITASNAVOGEM
 TTVIKTMDIADSTORIGTITSLINDIAFTQNTIALNAEAAAGAGOGGFAVABE
 VRHLASANAANDIRKLIDASADKVOGSGOQVIAARTMEDIVAAQVNTQTLAQS
 HSTLEQADGLSILTRAVDELNLITQKNAELVEEAGVSAMVKHRSLEDAVYTLH"
 2778..4268

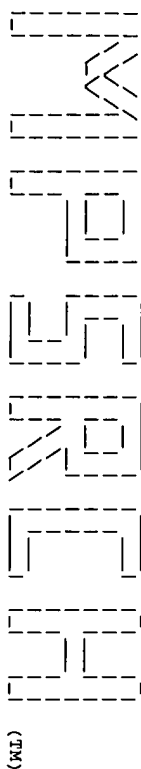
CDS

/gene="y9j6"
 2778..4268
 /gene="y9j6"
 /Ec_number="2.6.1.13"
 /note="0496: 100 pct identical to 429 amino acids of

misc_structure

CDS
12418..14436

/note= "predicted bend of 78..75 degrees"



(TM)

Release 3.0.4A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1997 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch.un n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Fri Apr 24 08:20:39 1998; Maspar time 349.28 Seconds

Tabular output not generated. 1215.222 Million cell updates/sec

Title: >US-08-483-534A-1
Description: (94-597) from US08483534A.seq (1 of 2)

Perfect Score: 504

N.A. Sequence: 94 GAGGAGGTCATCCATCCCG.....TGAAAGGGGGGACATTAGC 597
Comp: CTCCTCCAGTAGGAGGAGGCG.....ACTTCCCTTGTAAATGC

Scoring table: TABLE default

Gap 6

Nmatch STR : Dbase 0; Query 0

Searched: 1126798 seqs, 421087984 bases x 2

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

emb1-est

1:em_est1 2:em_est2 3:em_est3 4:em_est4 5:em_est5

6:em_est6 7:em_est7 8:em_est8

9:em_est9 10:em_est10 11:em_est11 12:em_est12 13:em_est13

14:em_est14 15:em_est15 16:em_est16 17:em_est17 18:em_est18

19:em_est19 20:em_est20 21:em_est21 22:em_est22 23:em_est23

Statistics: Mean 10.346; Variance 1.999; scale 5.176

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query	Match	Length	DB	ID	Description	Pred.	No.
	1	438	86.9	475	19	AA152266	2104c02.r1	Soares preg	0.00e+00	
	2	402	79.8	533	11	H10555	ym04e04.r1	Homo sapien	0.00e+00	
	3	289	57.3	407	8	R18984	yg25d10.r1	Homo sapien	0.00e+00	
	4	278	55.2	459	15	AA102052	2178g12.r1	Stratagene	0.00e+00	
	5	275	54.6	525	15	AA086129	2184e05.r1	Stratagene	0.00e+00	
	6	267	53.0	337	17	AA154293	mq49e07.r1	Soares 2NDM	0.00e+00	
	7	219	43.5	243	14	N87033	LI914F	Fetal heart, La	0.00e+00	
	8	215	42.7	423	14	AA003398	mg57b09.r1	Soares mous	0.00e+00	
	9	202	40.1	517	17	AA221081	mv65f03.r1	Soares mous	0.00e+00	
	10	193	38.3	248	18	AA327316	EST30628	Colon I Homo	0.00e+00	
	11	163	32.3	209	11	H22113	Y134g06.r1	Homo sapien	7.75e-276	
	12	157	31.2	500	14	AA049626	mj36d06.r1	Soares mous	3.33e-263	
	13	157	31.2	510	14	AA051121	mg74c08.r1	Soares mous	3.33e-263	
	14	113	22.4	437	11	H13315	Y172b07.r1	Homo sapien	8.64e-172	
	15	104	20.6	470	15	AA102053	2178g12.s1	Stratagene	1.89e-153	

C	16	100	19.8	308	19	AA369979	EST81510	Prostate gland	2.37e-145
	17	100	19.8	503	17	AA230715	mw04g08.r1	Soares mous	2.37e-145
	18	78	15.5	362	23	RI002054A	Rice cDNA, partial seq	1.49e-101	
	19	73	14.5	587	15	AA120159	mn33c11.r1	Beddington	8.03e-92
	20	73	14.5	691	17	AA218315	mw55a01.r1	Soares mous	8.03e-92
	21	69	13.7	578	17	AA138243	mq78g03.r1	Stratagene	4.12e-84
	22	67	13.3	522	18	AA289897	vb14c06.r1	Soares mous	2.76e-80
	23	65	12.9	438	16	AA198745	mw39a01.r1	Soares mous	1.77e-76
	24	64	12.7	506	14	W30441	mc21c12.r1	Soares mous	1.39e-74
	25	64	12.7	508	17	AA242568	mx24e07.r1	Soares mous	1.39e-74
	26	56	11.1	338	22	AA435309	ve15d08.r1	Soares mous	1.33e-59
	27	53	10.5	193	12	N50217	yy79f05.r1	Homo sapien	3.83e-54
	28	53	10.5	441	17	AA268851	va44c02.r1	Soares mous	3.83e-54
	29	50	9.9	619	15	W90821	mf78d07.r1	Soares mous	1.01e-48
	30	46	9.1	360	22	C13934	C.elegans cDNA clone Y	1.26e-41	
	31	46	9.1	411	14	AA033217	mi37d12.r1	Soares mous	1.26e-41
	32	44	8.7	420	14	W55273	me33d03.r1	Soares mous	3.88e-38
	33	44	8.7	480	14	W57595	me51g01.r1	Soares mous	3.88e-38
	34	44	8.7	495	14	W55249	me53c04.r1	Soares mous	3.88e-38
	35	43	8.5	427	23	AA456374	aa14e06.r1	Soares NHM	2.06e-36
	36	43	8.5	454	23	AA456533	aa15b05.r1	Soares NHM	2.06e-36
	37	37	7.3	428	16	AA173393	zp47c10.r1	Stratagene	2.48e-26
	38	37	7.3	428	15	AA134849	zn90e06.r1	Stratagene	2.48e-26
C	39	35	6.9	475	23	AA455165	aa15b05.s1	Soares NHM	4.28e-23
C	40	35	6.9	487	23	AA455908	aa14e06.s1	Soares NHM	4.28e-23
	41	33	6.5	295	22	C03071	Human Heart cDNA, clone	6.23e-20	
	42	33	6.5	317	22	C03104	Human Heart cDNA, clone	6.23e-20	
	43	29	5.8	183	15	AA090646	Y1128.seq.F	Fetal heart	7.31e-14
	44	28	5.6	428	19	AA150204	2104c02.s1	Soares preg	2.09e-12
	45	27	5.4	141	16	AA064500	ml50e08.r1	Stratagene	5.57e-11

ALIGNMENTS

RESULT	1
LOCUS	AA152266 475 bp mRNA 14-MAY-1997
DEFINITION	2104c02.r1 Soares pregnant uterus NBHPV Homo sapiens cDNA clone
	491330.5' similar to TR:G498912 G498912 ENDOTHELIAL-MONOCYTE
	ACTIVATING POLYPEPTIDE II. ;
ACCESSION	AA152266
NID	g1721603
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
	Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata;
	Vertebrata; Mammalia; Euthera; Primates; Catarrhini; Homidae;
	Homo.
REFERENCE	1 (bases 1 to 475)
AUTHORS	Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
	Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
	Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Treviski,E.,
	Waterson,R., Williamson,A., Wohlmann,P. and Wilson,R.
	WashU-Merck EST project
TITLE	Unpublished (1995)
JOURNAL	
COMMENT	Contact: Wilson RK
	WashU-Merck EST project
	Washington University School of Medicine
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
	Tel: 314 286 1800
	Fax: 314 286 1810
	Email: est@wustl.wustl.edu
	This clone is available royalty-free through LMT; contact the
	IMAGE Consortium (infoimage.lln.gov) for further information.
	Insert Length: 1171 Std Error: 0.00
	Seq primer: -28M13 rev2 from Amerisham
	High quality sequence stop: 431.
FEATURES	Location/Qualifiers
SOURCE	1..475
	/organism="Homo sapiens"
	/note="Organ: uterus; Vector: pT7T3-Pac; Site_1: Not I;
	Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
	oligo(dT) primer [5'

rs1e2-Hind III whole brain from a 73 days post natal female. 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', AACTGAGAAATTCGCGCCGAGGAAATTTTCTTTTCTT 3']; double-stranded cDNA was ligated to Hind III adaptors (Pharmacia), digested with Not I and directionally cloned into the Not I and Hind III sites of the Laimd BA vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Patima Bonaldi.

ORGANISM Homo sapiens
Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 407)
Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Treviskis,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.

TITLE The Washu-Merc EST Project
JOURNAL Unpublished (1995)
COMMENT GDB: G00-406-036
Contact: Wilson RK
Washu-Merc EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
High quality sequence stops: 294
Source: IMAGE Consortium, LMLU
This clone is available royalty-free through LML; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

FEATURES
source 1.407
/organism="Homo sapiens"
/clone="33689"

BASE COUNT 103 a 87 c 121 g 95 t 1 others
ORIGIN

Query Match 57.3%; Score 289; DB 10; Length 407;
Best Local Similarity 94.5%; Pred. No. 0.00e+00;
Matches 364; Conservative 0; Mismatches 10; Indels 11; Gaps 11;

Db 1 AAGATTGACGTGGGGAGCTGAACCAAGAGCTGTGTGAGCGGCTGTGACAGTTGCTG 60
|||
Oy 184 AAGATTGACGTGGGGAGCTGAACCAAGAGCTGTGTGAGCGGCTGTGACAGTTGCTG 243
|||
Db 61 CCCAAGAGAGAACTGCGAGACAGGCTGTGTGCTGTGCAACTGAAACCCCAAG 120
|||
Oy 244 CCCAAGAGAGAACTGCGAGACAGGCTGTGTGCTGTGCAACTGAAACCCCAAG 303
|||
Db 121 ATGAGAGAGCGAGTCCCAAGGATGCTGTGTGCTGTATAGAAAGATAAACCGC 180
|||
Oy 304 ATGAGAGAGCGAGTCCCAAGGATGCTGTGTGCTGTATAGAAAGATAAACCGC 363
|||
Db 181 CAGTTGAACTTGTGACCTCCGCGAGGCTTGTCTCTGCGTGAGCAAGCTTTGTG 240
|||
Oy 364 CAGTTGAACTTGTGACCTCCGCGAGGCTTGTCTCTGCGTGAGCAAGCTTTGTG 420
|||
Db 241 AAGGGCTATGAAAGGGCCCAACGATGAGAGCTCAAGCCCAAGAAAGTCTTGA 300
|||
Oy 421 AAGGGCTATGAAAGGGCCCAACGATGAGAGCTCAAGCCCAAGAAAGTCTTGA 479
|||
Db 301 GAAGTTGCGAGGCTGTGCTTCAAAATTTCTGTGAGAGTGCATCCGAAGTGGGAACA 360
|||
Oy 480 GAAGTTT-GCAGGCTT-GACTT-CAAAATTTCTT-GAGGAGTGAT-CGACACAGTGG-AAACA 533
|||
Db 361 AACCACTTTCATGACCAAGTTGGG 385
|||
Oy 534 AACCACTTTCAT-GACCAAGTGGG 557
|||

RESULT 4

LOCUS AA102052 459 bp mRNA EST 28-OCT-1996
DEFINITION 2178412.r1 Strataene colon (#937204) Homo sapiens cDNA clone
510790 5' similar to WP:F58B3.5 CE06007 METHIONYL-TRNA SYNTHETASE

ACCESSION AA102052
NID 91645892
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eucaryotae; mitochondria eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 459)
Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Treviskis,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.

TITLE Washu-Merc EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
Washu-Merc EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LML; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28W13 rev2 from Amersham
High quality sequence stop: 418.

FEATURES
source 1.459
/organism="Homo sapiens"
/note="Organ: colon; Vector: pBluescript SK-; Site:1: EcoRI; Site:2: XhoI; Cloned unidirectionally. Primer: Oligo dT. T-84 colonic epithelial cell line. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTTCTT 3'."
/db_xref="taxon:9606"
/clone="510790"
/lab_host="SOLR cells (kanamycin resistant)"
/ab_host="SOLR cells (kanamycin resistant)"
/ab_host="SOLR cells (kanamycin resistant)"

BASE COUNT 114 a 118 c 117 g 107 t 3 others
ORIGIN

Query Match 55.2%; Score 278; DB 15; Length 459;
Best Local Similarity 95.6%; Pred. No. 0.00e+00;
Matches 368; Conservative 0; Mismatches 2; Indels 15; Gaps 15;

Db 1 AAGGGCTGTGTACAGTTGCTGCCAAGAGAGAACTGACAGAGCTGTGTGTGCTG 60
|||
Oy 223 AAGGGCTGTGTACAGTTGCTGCCAAGAGAGAACTGACAGAGCTGTGTGTGCTG 282
|||
Db 61 TGCACCTGAAACCCCAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
|||
Oy 283 TGCACCTGAAACCCCAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 342
|||
Db 121 TCTATGAAAGGATTAACCCAG-TTGAACCTTGTGACCTCCGAG-CTCTGCTCT 178
|||
Oy 343 TCTATGAAAGGATTAACCCAG-TTGAACCTTGTGACCTCCGAG-CTCTGCTCT 402
|||
Db 179 GGTGAGCAGCGTGTGTGAGGCTATGAAAGGCAACAGATGAGAGCTCAAGCC 238
|||
Oy 403 GGTGAGCAG-CTGTGTGTGAGGCTATGAAAGGCAACAGATGAGAGCTCAAGCC 461
|||
Db 239 CAGAGAAAGGCTTGTGAGAAAGTGCAGGCTCAAAATTTCTTGAAGAGAGCAT 298
|||
Oy 462 CAGAGAAAGGCTTGTGAGAA-CTTGCAGGCTGACTTCAAAATTTCTT-GA-GGAGTGCAT 518
|||
Db 299 CACCAAGTTGGAACCAACCTTCAATTGACCAAGCTTGCGCTCAATTCTCTGTTA 358
|||

OY 519 GCGACA-GT-GGAAGCAAAACCACTT-CAT-GACCAAGCT-GGGCTCCATTTC-TGT-A 571
 DB 359 AATCGCTGAAAAGGGGGAACATT 383
 OY 572 AATCGCT-GAAA-GGGGGGGAACATT 594
 RESULT 5
 LOCUS AA086129 525 bp mRNA EST 23-OCT-1996
 DEFINITION 2184e05.r1 Stratiogene colon (#937204) Homo sapiens cDNA clone
 511328 5' similar to TR:G1184659 G1184659 TYROSL-TRNA SYNTHETASE.
 ACCESSION AA086129
 NID g1629697
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata;
 Vertebrata; Eutheria; Primates; Catarrhini; Homindae; Homo.
 REFERENCE 1 (bases 1 to 525)
 Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
 Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevisan, E.,
 Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
 WashU-Werck EST Project
 Unpublished (1995)
 TITLE JOURNAL
 COMMENT
 Contact: Wilson RK
 WashU-Werck EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -28M13 rev2 from Amersham
 High quality sequence stop: 330.
 Location/Qualifiers
 1..525
 /organism="Homo sapiens"
 /note="Organ: colon; Vector: pBluescript SK-; Site: 1;
 EcoRI; Site: 2: XhoI; Cloned unidirectionally. Primer:
 Oligo dt. T-84 colonic epithelial cell line. Average
 insert size: 1.0 kb; Uni-ZAP XR Vector: -5' adaptor
 sequence: 5' GAATTCGGACGACG 3' -3' adaptor sequence: 5'
 CTCGAGTCTTTTCTTTTCTTTT 3'."
 /db_xref="taxon:9606"
 /clone="511328"
 /clone_lib="Stratiogene colon (#937204)"
 /lab_host="SOLR cells (kanamycin resistant)"
 <1..>525
 BASE COUNT 144 a 125 c 149 g 105 t 2 others
 ORIGIN
 mRNA
 Query Match 54.6%; Score 275; DB 15; Length 525;
 Best Local Similarity 96.4%; Pred. No. 0.00e+00;
 Matches 322; Conservative 0; Mismatches 5; Indels 7; Gaps 7;
 DB 155 GAGGAGGTATCCATCCCGGCTGGATATCCGTTGGGAAAATCATCACTGTGGAAG 214
 OY 94 GAGGAGGTATCCATCCCGGCTGGATATCCGTTGGGAAAATCATCACTGTGGAAG 153
 DB 215 CACCCAGATGACAGACGCTGTATGAGAGAGATGAGTGGGGGAACCTGAACACGG 274
 OY 154 CACCCAGATGACAGACGCTGTATGAGAGAGATGAGTGGGGGAACCTGAACACGG 213
 DB 275 ACTGTGCTGACAGCGCTGTACAGTCTGCGCCAGAGGAATTCAGAGACGCTGCT 334
 OY 214 ACTGTGCTGACAGCGCTGTACAGTCTGCGCCAGAGGAATTCAGAGACGCTGCT 272
 DB 335 AGTGTGCTGCTCAACCTGAACCCAGAGATGAGAGGAGTGAAGTCCCAAGGCATC 394
 OY 335 AGTGTGCTGCTCAACCTGAACCCAGAGATGAGAGGAGTGAAGTCCCAAGGCATC 394

OY 273 AGTGTGCTGCTCAACCTGAACCCAGAGATGAGAGG-AGTCAAGTCCCAAGGCATC 331
 DB 395 TTCTGTGCTGCTCTATGAGAGGATTAACCGCCAGTTGAACCTTGACCCCGCCG 454
 OY 332 TTCTGTGCTGCTCTATGAGAGGATTAACCGCCAGTTGAACCTTG-AGACCTTCGCCG 389
 DB 455 AAGGCTGTGCTCCCTGTGAGCAGCTGTTTGG 488
 OY 390 A-GCGCTGCTCTCC-TGCT-GACCAAGTGTGTTGTG 420
 RESULT 6
 LOCUS AA154293 337 bp mRNA EST 19-FEB-1997
 DEFINITION mq49e07.r1 Soares 2NbMT Mus musculus cDNA clone 582084 5' similar
 to TR:G498912 G498912 ENDOTHELIAL-MONOCYTE ACTIVATING POLYPEPTIDE
 II.
 ACCESSION AA154293
 NID g1725961
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata;
 Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Mus.
 REFERENCE 1 (bases 1 to 337)
 Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
 Gelsel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
 Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
 Theising, R., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
 Waterston, R.
 The WashU-HMI Mouse EST Project
 Unpublished (1996)
 TITLE JOURNAL
 COMMENT
 Contact: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -28M13 rev2 from Amersham
 High quality sequence stop: 324.
 Location/Qualifiers
 1..337
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site: 1: Not I; Site: 2: Eco RI; 1st strand cDNA
 was primed with a Not I - oligo(dt) primer [5',
 TGTTACCAATCTGAAGTGGGAGCGCGCGTGTGTTTTTTTTTTTTTTTTTTT
 3']; double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT73 vector. RNA
 provided by Dr. Bertrand Jordan. Library went through two
 rounds of normalization, and was constructed by Bento
 Soares and M. Fatima Bonaldo."
 /db_xref="taxon:10090"
 /clone="582084"
 /clone_lib="Soares 2NbMT"
 /db_xref="582084"
 /tissue_type="Thymus"
 /dev_stage="4 weeks"
 /lab_host="DH10B"
 <1..>337
 BASE COUNT 82 a 74 c 113 g 68 t
 ORIGIN
 mRNA
 Query Match 53.0%; Score 267; DB 17; Length 337;
 Best Local Similarity 89.6%; Pred. No. 0.00e+00;
 Matches 302; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

```

Db 1 GCACCCAGATGCATGACCTTTATGTGGAGAGATGTGTGGGGAGACTGAACCCAG 60
QY 153 GCACCCAGATGCATGACCTTTATGTGGAGAGATGTGTGGGGAGACTGAACCCAG 212
Db 61 GACGTGGTGAAGCGGCGCTGTCACATTTGTGCCCAAGAAAGACATGCAGACAGCTGCT 120
QY 213 GACGTGGTGAAGCGGCGCTGTCACATTTGTGCCCAAGAAAGACATGCAGACAGCTGCT 272
Db 121 GGTGGTGTGTGCAATCTGAACCCAGAGATGAGAGCGCTGACCTGCGAGCGATGCT 180
QY 273 AGTGGTGTGTGCAATCTGAACCCAGAGATGAGAGCGCTGACCTGCGAGCGATGCT 332
Db 181 ACTGTGCTCTCTGTAGAAGGGGTAGCCGCAAGTTGAACCTCTGACCCCTCTGCTGG 240
QY 333 TCTGTGCTCTCTATATAGAAGGATTAACCGCAGGTGTAACCTCTGACCCCTCTGCGAGG 392
Db 241 CTCGTCTCTCTGTGACAGAGTGTGTGACAGGCTATGAGAAAGGCCAGCCAGATGAGA 300
QY 393 CTCGTCTCTCTGTGACAGAGTGTGTGACAGGCTATGAGAAAGGCCAGCCAGATGAGA 452
Db 301 GCTCAAGCCCAAGAGAAAGTCTTGTAGAAGCTGCAG 337
QY 453 GCTCAAGCCCAAGAGAAAGTCTTGTAGAAGCTGCAG 489

RESULT 7
LOCUS N87033 243 bp mRNA EST 01-APR-1996
DEFINITION L1914F fetal heart, lambda ZAP Express Homo sapiens cDNA clone
L1914.5' similar to EST(Y625D10.R1 F23395 METHIONYL-TRNA
SYNTHETASE).
ACCESSION N87033
NID g1440235
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Euthera; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 243)
Llew.C.C.
CDNAs from fetal heart (1996)
JOURNAL Unpublished (1996)
COMMENT
Contact: Llew CC
Molecular Cardiology
University of Toronto
Banting Institute, 100 College St., Toronto, Ontario, M5G1L5
Tel: 4169788758
Fax: 4169785650
Email: llewcc@utcc.utoronto.ca
Seq primer: GAATTTAACTCTCACTAAAGC.
FEATURES
source
1.243
/organism="Homo sapiens"
/Note="Vector: Lambda ZAP Express; Site_1: EcoRI; Site_2:
XhoI; mRNA was purified from human fetal hearts (8-10
weeks). cDNA was synthesized using a XhoI-oligo dT
adaptor-primer. EcoRI adaptors were ligated, followed by
digestion with XhoI for directional cloning into
preligated lambda ZAP Express."
/db_xref="taxon:9606"
/clone_id="L1914"
/clone_1db="Fetal heart, Lambda ZAP Express"
/lab_host="E. coli XL1-Blue"
<1..>243
BASE COUNT 57 a 58 c 77 g 51 t
ORIGIN
mRNA
57 a 58 c 77 g 51 t
Query Match 43.5%; Score 219; DB 14; Length 243;
Best Local Similarity 96.7%; Pred. No. 0.00e+00;
Matches 237; Conservative 0; Mismatches 6; Indels 2; Gaps 2;
Db 1 CCCAGATGCACAGACCTGTATGTAGAGAGATTGACGTGGGGAGCTGAACACGAGC 60

```

```

QY 156 CCGAGATGCACAGACCCCTGTATGTAGAGAAAGATTGACGTGGGGAGCTGAACACGAGC 215
Db 61 TGTGGTGAAGCGGCGCTGTCACATTTGTGCCCAAGAAAGACATGCAGACAGCTGCT 120
QY 216 TGTGGTGAAGCGGCGCTGTCACATTTGTGCCCAAGAAAGACATGCAGACAGCTGCT 275
Db 121 TGTGGTGTGTGCAATCTGAACCCAGAGATGAGAGCGCTGACCTGCGAGCGATGCT 179
QY 276 GGTGGTGTGTGCAATCTGAACCCAGAGATGAGAGCGCTGACCTGCGAGCGATGCT 335
Db 180 GTGTCTCTCTATATAGAAGGATTA-CCGCGCAGTTGAACCTCTTGGCCCTCGGAGGTTTC 238
QY 336 GTGTCTCTCTATATAGAAGGATTAACCGCAGGTGTAACCTCTGACCCCTCTGCGAGGCTC 395
Db 239 TGCCTC 243
QY 396 TGCCTC 400

RESULT 8
LOCUS AA003398 423 bp mRNA EST 19-JUL-1996
DEFINITION ms57b09.t1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
clone 437081 5' similar to PIR:A55053 A55053
endothelial monocyte-activating protein II precursor - mouse ;.
ACCESSION AA003398
NID g1446863
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Euthera; Rodentia; Sciurognathi; Myomorpha; Muridae;
Murinae; Mus.
1 (bases 1 to 423)
Marra,M., Hallier,J., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Reising,B., Wyllie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LNC; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:262417
Seq primer: ETPprimer
High quality sequence stop: 343.
FEATURES
source
1.423
/organism="Mus musculus"
/strain="C57BL/6J"
/Note="Vector: p773D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTACCAATCTGACAGGAGGCGGCGGAGAAATTTTGTGTGTGTGTGTGTGTGT
T 3']; on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
State Univ., from 2 j; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
p773 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Fatima Bonaldo."
/db_xref="taxon:10090"
/clone_id="437081"
/clone_1db="Soares mouse embryo NbME13.5 14.5"

```

	ORIGIN	BASE COUNT	mrna	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521
--	--------	------------	------	---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

FEATURES	source
LOCUS	AA221081 517 bp
DEFINITION	musf03.r1 Soares mouse 3NNE12.5 Mus musculus cDNA clone 659933 5'
ACCESSION	AA221081
NID	AA221081
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryotes: mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	1 (bases 1 to 517) Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE	The WashU-HHMI Mouse EST Project
JOURNAL	Unpublished (1996)
COMMENT	Contact: Marra M/Mouse EST Project WashU-HHMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel.: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.lnl.gov) for further information. MGI:405781 Seq primer: -28m13 rev2 ET from Amersham High quality sequence stop: 493. Location/Qualifiers 1..517

```

FEATURES
source
    Location/Qualifiers
        1..517
            /organism="Mus musculus"
            /strain="C57BL/6J"
            /note="vector: pRT3-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5]

```

[illegible]

REFERENCE	10	EST	20-APR-1997
LOCUS	AA327316	248 bp	mRNA
DEFINITION	EST30628	Colon I Homo sapiens	CDNA 5' end.
ACCESSION	AA327316		
NID	g1979561		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryotes: mitochondrial eukaryotes; Metazoa; Chordata;		
	Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae;		
	Homo.		
	1 (bases 1 to 248)		
ADAMS,M.D., Kerlavage,A.R., Fletschmann,R.D., Fuldner,R.A.,			
Bull,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,			
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wei,C.,			
Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,			
Fitzhugh,L.M., Fitzhugh,W.M., Fritchman,J.L., Geognagen,N.S.,			
Glocke,A., Gnelum,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,			
Kelley,J.M., Kelley,C.J., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,			
Moreno-balanquas,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,			
Phillips,C.A., Ryder,S.E., Scott,L.J., Saudek,D.M., Shirley,R.,			
Small,K.V., Springs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,			
Bedarick,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,			
Dimke,D., Feng,D.-F., Ferrite,A., Fischer,C., Hastings,E.J.,			
H.W.W., Hu,J.-S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,			
Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H.,			
Raymond,L., Wei,Y.-F., Wang,J., Xu,C., Ye,G.-L., Ruben,S.M.,			
Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,			
Fraser,C.M. and Venter,J.C.			
Initial assessment of human gene diversity and expression patterns			
based upon 83 million nucleotides of cDNA sequence			

FEATURES

Location/Qualifiers
1. 500

/organism="Mus musculus"

/strain="C57BL/6J"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA

was primed with a Not I - oligo(dT) primer [5',

TGTACCAATCTGAGTGGAGCGCGCGGGAATTTTCTTTTCTTTT

T 3'], on equal amounts of mRNA from 2 13.5dpc and 2

14.5dpc embryos (total RNA provided by Minoru Ko, Wayne

State Univ., from 2]; double-stranded cDNA was ligated to

Eco RI adaptors (Pharmacia), digested with Not I and

cloned into the Not I and Eco RI sites of the modified

pT73 vector. Library went through one round of

normalization, and was constructed by Bento Soares and

M. Fatima Bonaldo."

/db_xref="taxon:10090"

/clone="478187"

/clone_lib="Soares mouse embryo NBM13.5 14.5"

/sex="unknown"

/tissue_type="embryo"

/dev_stage="13.5-14.5dpc total fetus"

/lab_host="DH10B"

<1. >500

BASE COUNT

130 a 133 c 130 g 107 t

ORIGIN

Query Match

Best Local Similarity 31.2%; Score 157; DB 14; Length 500;

Matches 177; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Db

1 CTGTGAAAGAGTCTTGTGACAGGCTATGAGAGGCGCCAGCAGATGAGAGCTCAAGC 60

QY 401 CTGTGAGCAGCTGTGTGTGAGAGGCTATGAGAGGCGCCAGCAGATGAGAGCTCAAGC 460

Db

61 CCAAGAGAAAGTCTTGTGAGAGCTGCGAGCGGCGACTTAAATTTCTGAGAGTGCATCG 120

QY 461 CCAAGAGAAAGTCTTGTGAGAGCTGCGAGCGGCGACTTAAATTTCTGAGAGTGCATCG 520

Db

121 CACAGTGAAGCAACCAACTTCATGACCAAGCTGGGATTCGTCCTGTAATCACTAA 180

QY 521 CACAGTGAAGCAACCAACTTCATGACCAAGCTGGGATTCGTCCTGTAATCACTAA 580

Db

181 AAGGGGGTAACTATCAGC 197

QY 581 AAGGGGGTAACTATCAGC 597

RESULT 13

LOCUS AA051121 510 bp mRNA EST 09-SEP-1996

DEFINITION mg74c08.r1 Soares mouse embryo NBM13.5 14.5 Mus musculus cDNA

clone 438734 5'.

ACCESSION AA051121

NID G1530793

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;

Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Mus.

1 (bases 1 to 510)

Marra,M., Hallier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,

Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,

Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,

Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and

Waterston,R.

The WashU-HMI Mouse EST Project

Unpublished (1996)

TITLE

JOURNAL

COMMENT

Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

FEATURES

source

Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LBNL; contact the
IMAGE Consortium (info@image.lbnl.gov) for further information.
MI:264070
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 495.
Location/Qualifiers
1. 510

/organism="Mus musculus"

/strain="C57BL/6J"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA

was primed with a Not I - oligo(dT) primer [5',

TGTACCAATCTGAGTGGAGCGCGCGGGAATTTTCTTTTCTTTT

T 3'], on equal amounts of mRNA from 2 13.5dpc and 2

14.5dpc embryos (total RNA provided by Minoru Ko, Wayne

State Univ., from 2]; double-stranded cDNA was ligated to

Eco RI adaptors (Pharmacia), digested with Not I and

cloned into the Not I and Eco RI sites of the modified

pT73 vector. Library went through one round of

normalization, and was constructed by Bento Soares and

M. Fatima Bonaldo."

/db_xref="taxon:10090"

/clone="438734"

/clone_lib="Soares mouse embryo NBM13.5 14.5"

/sex="unknown"

/tissue_type="embryo"

/dev_stage="13.5-14.5dpc total fetus"

/lab_host="DH10B"

<1. >510

BASE COUNT

134 a 132 c 130 g 114 t

ORIGIN

Query Match

Best Local Similarity 31.2%; Score 157; DB 14; Length 510;

Matches 177; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Db

1 CTGTGAAAGAGTCTTGTGACAGGCTATGAGAGGCGCCAGCAGATGAGAGCTCAAGC 60

QY 401 CTGTGAGCAGCTGTGTGTGAGAGGCTATGAGAGGCGCCAGCAGATGAGAGCTCAAGC 460

Db

61 CCAAGAGAAAGTCTTGTGAGAGCTGCGAGCGGCGACTTAAATTTCTGAGAGTGCATCG 120

QY 461 CCAAGAGAAAGTCTTGTGAGAGCTGCGAGCGGCGACTTAAATTTCTGAGAGTGCATCG 520

Db

121 CACAGTGAAGCAACCAACTTCATGACCAAGCTGGGATTCGTCCTGTAATCACTAA 180

QY 521 CACAGTGAAGCAACCAACTTCATGACCAAGCTGGGATTCGTCCTGTAATCACTAA 580

Db

181 AAGGGGGTAACTATCAGC 197

QY 581 AAGGGGGTAACTATCAGC 597

RESULT 14

LOCUS H13315 437 bp mRNA EST 27-JUN-1995

DEFINITION y172007.r1 Homo sapiens cDNA clone 43366 5' similar to

SP:SYQC_YEAST P36421 TYROSYL-TRNA SYNTHETASE, CYTOPLASMIC ;

H13315

ACCESSION 9878135

NID

KEYWORDS EST.

SOURCE human clone=43366 library=Soares infant brain INB vector=Lambda

host=DH10B (ampicillin resistant) primer=H13P1 RstIel-Not I

RstIel2-Hind III Whole brain from a 73 days post natal female. 1st

strand cDNA was primed with a Not I - oligo(dT) primer [5',

AACTGGAAGATTCGCGCGGCGGAGGAGTATTTTCTTTTCTTTT

T 3'], double-stranded

cDNA was ligated to Hind III adaptors (Pharmacia), digested with

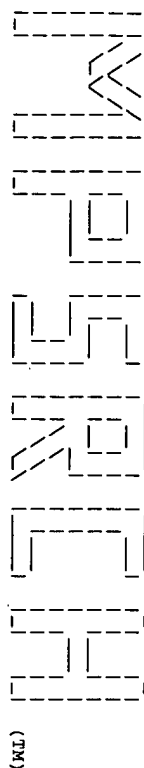
Not I and directionally cloned into the Not I and Hind III sites of

the Lambda BA vector. Library went through one round of

normalization. Library constructed by Bento Soares and M. Fatima

Bonaldo.

CONTACT: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108


 (TM)

Release 3.0.4A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1997 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Mparch_n n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Fri Apr 24 08:13:01 1998; Maspar time 430.29 Seconds
 1384.997 Million cell updates/sec

Tabular output not generated.

Title: >US-08-483-534A-1
 Description: (94-597) from US08483534A.seq (1 of 2)
 Perfect Score: 504
 N.A. Sequence: 94 GAGGAGGTGATCCCATCCG.....TGAAGGGGGGACATTAGC 597
 Comp: CTCCTCCAGTAGGTTAGGCG.....ACTTCCCTTCCTTATTCG

Scoring table: TABLE default
 Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 354530 segs, 591221332 bases x 2

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: emb153

1:em_in 2:em_or 3:em_com 4:em_ov 5:em_pl 6:em_htg
 7:em_hum1 8:em_hum2 9:em_da 10:em_ro 11:em_un 12:em_vi
 13:em_pat
 genbank105
 14:gb_ro 15:gb_com 16:gb_ov 17:gb_in 18:gb_pl 19:gb_da
 20:gb_st 21:gb_vi 22:gb_ph 23:gb_sy 24:gb_un 25:gb_pat
 26:gb_htg

Statistics: Mean 10.143; Variance 5.076; scale 1.998

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
C 1	82	16.3	21948	17	CEFS8B3	Caenorhabditis elegans 9.54e-41
C 2	82	16.3	85953	26	CEH04L24	Caenorhabditis elegans 9.54e-41
C 3	82	16.3	131318	26	CEY42E10	Caenorhabditis elegans 9.54e-41
C 4	81	16.1	498	23	ASEMAP11	Artificial sequence DN 5.56e-40
C 5	73	14.5	1068	14	NMDU10118	Mus musculus endoderm 6.67e-34
C 6	73	14.5	1086	25	150092	Sequence 35 from paten 6.67e-34
C 7	72	14.3	7218	25	166494	Sequence 14 from paten 6.67e-34
C 8	66	13.1	1254	14	AF021800	Cricketulius griseus mul 1.15e-28
C 9	33	6.5	1703	18	SCYGL105W	S.cerevisiae chromosome 9.53e-06
C 10	33	6.5	2473	24	SCARCI	S.cerevisiae ARCI gene 9.53e-06
C 11	33	6.5	12325	18	SCVIL1EFT	S.cerevisiae fragment 9.53e-06
C 12	32	6.3	1745	18	SCU31348	Saccharomyces cerevisi 3.93e-05
C 13	31	6.2	215	25	128278	Sequence 5 from patent 1.59e-04
C 14	30	6.0	11157	19	ECAE000389	Escherichia coli K-12 6.34e-04

RESULT	LOCUS	DEFINITION	ACCESSION	NID	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT		
1	CEFS8B3	21948 bp	DNA	INV	23-JAN-1998	Caenorhabditis elegans	cosmid F58B3, complete sequence.	1	(bases 1 to 21948)	1	Escherichia coli K-12	6.34e-04	1	Archaeoglobus fulgidus	1.28e-01	1	Drosophila melanogaster	4.52e-01
2	CEH04L24	85953 bp	DNA	INV	23-JAN-1998	Caenorhabditis elegans	cosmid F58B3, complete sequence.	2	(bases 1 to 21948)	2	Drosophila melanogaster	4.52e-01	2	Drosophila melanogaster	4.52e-01	2	Drosophila melanogaster	4.52e-01
3	CEY42E10	131318 bp	DNA	INV	23-JAN-1998	Caenorhabditis elegans	cosmid F58B3, complete sequence.	3	(bases 1 to 21948)	3	Drosophila melanogaster	4.52e-01	3	Drosophila melanogaster	4.52e-01	3	Drosophila melanogaster	4.52e-01
4	ASEMAP11	498 bp	DNA	INV	23-JAN-1998	Caenorhabditis elegans	cosmid F58B3, complete sequence.	4	(bases 1 to 21948)	4	Drosophila melanogaster	4.52e-01	4	Drosophila melanogaster	4.52e-01	4	Drosophila melanogaster	4.52e-01
5	NMDU10118	1068 bp	DNA	INV	23-JAN-1998	Caenorhabditis elegans	cosmid F58B3, complete sequence.	5	(bases 1 to 21948)	5	Drosophila melanogaster	4.52e-01	5	Drosophila melanogaster	4.52e-01	5	Drosophila melanogaster	4.52e-01
6	150092	1086 bp	DNA	INV	23-JAN-1998	Caenorhabditis elegans	cosmid F58B3, complete sequence.	6	(bases 1 to 21948)	6	Drosophila melanogaster	4.52e-01	6	Drosophila melanogaster	4.52e-01	6	Drosophila melanogaster	4.52e-01
7	166494	7218 bp	DNA	INV	23-JAN-1998	Caenorhabditis elegans	cosmid F58B3, complete sequence.	7	(bases 1 to 21948)	7	Drosophila melanogaster	4.52e-01	7	Drosophila melanogaster	4.52e-01	7	Drosophila melanogaster	4.52e-01
8	AF021800	1254 bp	DNA	INV	23-JAN-1998	Caenorhabditis elegans	cosmid F58B3, complete sequence.	8	(bases 1 to 21948)	8	Drosophila melanogaster	4.52e-01	8	Drosophila melanogaster	4.52e-01	8	Drosophila melanogaster	4.52e-01
9	SCYGL105W	1703 bp	DNA	INV	23-JAN-1998	Caenorhabditis elegans	cosmid F58B3, complete sequence.	9	(bases 1 to 21948)	9	Drosophila melanogaster	4.52e-01	9	Drosophila melanogaster	4.52e-01	9	Drosophila melanogaster	4.52e-01
10	SCARCI	2473 bp	DNA	INV	23-JAN-1998	Caenorhabditis elegans	cosmid F58B3, complete sequence.	10	(bases 1 to 21948)	10	Drosophila melanogaster	4.52e-01	10	Drosophila melanogaster	4.52e-01	10	Drosophila melanogaster	4.52e-01
11	SCVIL1EFT	12325 bp	DNA	INV	23-JAN-1998	Caenorhabditis elegans	cosmid F58B3, complete sequence.	11	(bases 1 to 21948)	11	Drosophila melanogaster	4.52e-01	11	Drosophila melanogaster	4.52e-01	11	Drosophila melanogaster	4.52e-01
12	SCU31348	1745 bp	DNA	INV	23-JAN-1998	Caenorhabditis elegans	cosmid F58B3, complete sequence.	12	(bases 1 to 21948)	12	Drosophila melanogaster	4.52e-01	12	Drosophila melanogaster	4.52e-01	12	Drosophila melanogaster	4.52e-01
13	128278	215 bp	DNA	INV	23-JAN-1998	Caenorhabditis elegans	cosmid F58B3, complete sequence.	13	(bases 1 to 21948)	13	Drosophila melanogaster	4.52e-01	13	Drosophila melanogaster	4.52e-01	13	Drosophila melanogaster	4.52e-01
14	ECAE000389	11157 bp	DNA	INV	23-JAN-1998	Caenorhabditis elegans	cosmid F58B3, complete sequence.	14	(bases 1 to 21948)	14	Drosophila melanogaster	4.52e-01	14	Drosophila melanogaster	4.52e-01	14	Drosophila melanogaster	4.52e-01

ALIGNMENTS

1
 CEFS8B3 21948 bp DNA INV 23-JAN-1998
 Caenorhabditis elegans cosmid F58B3, complete sequence.
 273437
 91370027
 HTG: Arabidopsis DNA-damage-repair D11 protein like;
 Interferon-related PC4 protein like; Methionyl-tRNA synthetase.
 Caenorhabditis elegans.
 Caenorhabditis elegans
 Eukaryotes: Metazoa, Nematoda; Secernentea; Rhabdita; Rhabditiidae;
 Rhabditiina; Rhabditiidae; Rhabditiidae; Peloderiinae; Caenorhabditis.
 1 (bases 1 to 21948)
 Wilson, R., Almscough, R., Anderson, K., Baynes, C., Berks, M.,
 Bonfield, J., Burton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,
 Fulton, L., Garner, A., Green, P., Hawkins, J., Hillier, L., Jier, M.,
 Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N.,
 Latreille, P., Lightning, J., Lloyd, C., McKirry, B., Mortimore, B.,
 O'Callaghan, M., Parsons, J., Percy, C., Rifen, L., Roop, A.,
 Saunders, D., Showkeen, R., Smaildon, N., Smith, A., Sonnenhammer, E.,
 Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaubin, M.,
 Vaughan, K., Waterston, R., Watson, A., Weissman, L.,
 Wilkinson-Sprat, V., and Wohlman, P.
 2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans
 Nature in press
 2 (bases 1 to 21948)
 Harris, B.
 Direct Submission
 Submitted (13-MAY-1996) Louis, MO 63110, USA. E-mail:
 jseesanger.ac.uk or twenemate@wustl.edu
 Current sequence finishing criteria for the C. elegans genome
 sequencing consortium are that all bases are either sequenced
 unambiguously on both strands, or on a single strand with both a
 dye primer and dye terminator reaction, from distinct subclones.
 Exceptions are indicated by an explicit note. IMPORTANT: This
 sequence is NOT necessarily the entire insert of clone F58B3. It
 may be shorter because we only sequence overlapping sections once.

Cp	344	GAGGACACAGAAGCATGCCCTGGGACTCGAATCCTCTTCATCTTTGGGCTTCAAGTTG	285
Db	44246	CANAACACAAGACAGAGCGGTCTTGCAATTGATGCAGTGAAGCTGACGAAGCTCA	44305
Cp	284	CACACACCACCTATCACAGCGCTGCTCCGAGTTCTCTTTGGGACGAACTGTACCAAGCGC	225
Db	44306	GAATCACTAGTTCCGGGGGCTACTCTCAACATGCATGTGTCACGTAAGAACATCT	44365
Cp	224	CTCACCAACAGTCCGTGGTTCACGCTTCCCCACGTAATCTTTCTTACATACAGGCTGTCT	165
Db	44366	GCATCAGGATGCTTCTCACTATGATGATTCGACCAACGCGCATATCGAG	44415
Cp	164	GCATCTGGGTGCTTCTCCACAGTATGATGATTTCCCACACGAGATATTCAG	115

RESULT	LOCUS	DEFINITION	ACCESSION	NID	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE
4	ASEMAP11	Artificial sequence DNA for cytokine EMAP11.	Y08737	91619591	EMAP11 gene. synthetic construct. artificial sequence.	EMAP11 gene. synthetic construct. artificial sequence.	Schluesener H.J., Seid K., Zhao Y. and Meyermann R.	1 (bases 1 to 498)	Schluesener H.J., Seid K., Zhao Y. and Meyermann R.	Localization of endothelial monocyte-activating polypeptide II (EMAP-II), a novel proinflammatory cytokine, to lesions of experimental autoimmune encephalomyelitis, neuritis and uveitis: expression by monocytes and activated microglial cells
JOURNAL MEDLINE	REFERENCE									
AUTHORS										
TITLE										
1	(bases 1 to 498)									
2	(bases 1 to 498)									
3	(bases 1 to 498)									
4	(bases 1 to 498)									
5	(bases 1 to 498)									
6	(bases 1 to 498)									
7	(bases 1 to 498)									
8	(bases 1 to 498)									
9	(bases 1 to 498)									
10	(bases 1 to 498)									
11	(bases 1 to 498)									
12	(bases 1 to 498)									
13	(bases 1 to 498)									
14	(bases 1 to 498)									
15	(bases 1 to 498)									
16	(bases 1 to 498)									
17	(bases 1 to 498)									
18	(bases 1 to 498)									
19	(bases 1 to 498)									
20	(bases 1 to 498)									
21	(bases 1 to 498)									
22	(bases 1 to 498)									
23	(bases 1 to 498)									
24	(bases 1 to 498)									
25	(bases 1 to 498)									
26	(bases 1 to 498)									
27	(bases 1 to 498)									
28	(bases 1 to 498)									
29	(bases 1 to 498)									
30	(bases 1 to 498)									
31	(bases 1 to 498)									
32	(bases 1 to 498)									
33	(bases 1 to 498)									
34	(bases 1 to 498)									
35	(bases 1 to 498)									
36	(bases 1 to 498)									
37	(bases 1 to 498)									
38	(bases 1 to 498)									
39	(bases 1 to 498)									
40	(bases 1 to 498)									
41	(bases 1 to 498)									
42	(bases 1 to 498)									
43	(bases 1 to 498)									
44	(bases 1 to 498)									
45	(bases 1 to 498)									
46	(bases 1 to 498)									
47	(bases 1 to 498)									
48	(bases 1 to 498)									
49	(bases 1 to 498)									
50	(bases 1 to 498)									
51	(bases 1 to 498)									

DB	201	GAACCCGGCGAAATGCGTGGCGTGTCTCAGGCAATGTTATGTCGC	251
cy	291	GAACCCCGAGATGAGAGAGTCGAGTCCCAAGGCGATGCTGTGTGTGC	341
RESULT	5		
LOCUS	MMU10118	1068 bp	mRNA
DEFINITION	Mus musculus endothelial monocyte activating polypeptide II mRNA,		
ACCESSION	U10118		
NID	9498911		
KEYWORDS			
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Rodentia; Sclurognathi; Myomorphia; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 1068) Kao,J., Houck,K., Fan,Y., Haehnel,I., Libutti,S.K., Kayton,M.L., Glickselt,T., Chabot,J., Noygrud,R., Greenberg,S., Kuang,W.-J., Leung,D.W., Hayward,J.R., Kistiel,W., Heath,M., Brett,J. and Stern,D.M.		
TITLE	Characterization of a novel tumor-derived cytokine.		
JOURNAL	Endothelial-monocyte activating polypeptide II		
MEDLINE	J Biol. Chem. 269 (40), 25106-25119 (1994)		
REFERENCE	95014290		
AUTHORS	2 (bases 1 to 1068) Kao,J., Fan,Y., Haehnel,I., Brett,J., Greenberg,S., Claus,M., Kayton,M., Houck,K., Kistiel,W., Sejelid,R., Burnier,J. and Stern,D.		
TITLE	A peptide derived from the amino terminus of endothelial-monocyte activating polypeptide II modulates mononuclear and polymorphonuclear leukocyte functions, defines an apparently novel cellular interaction site, and induces an acute inflammatory response		
JOURNAL	J Biol. Chem. 269 (13), 9774-9782 (1994)		
MEDLINE	94193665		
REFERENCE	3 (bases 1 to 1068) Kao,J., Ryan,J., Brett,G., Chen,J., Shen,H., Fan,Y., Godman,G., Familletti,P.C., Wang,F., Pan,Y.E., Stern,D. and Claus,M.		
AUTHORS	Endothelial monocyte-activating polypeptide II. A novel tumor-derived polypeptide that activates host-response mechanisms		
TITLE	J Biol. Chem. 267 (28), 20239-20247 (1992)		
JOURNAL	93013897		
MEDLINE	4 (bases 1 to 1068) Houck,K.A.		
REFERENCE	Direct Submission		
AUTHORS	Submitted (27-MAY-1994) Keith A. Houck, Molecular Biology, Sphinx Pharmaceuticals Corp., P.O. Box 52330, Durham, NC 27711, USA		
JOURNAL	Pharmaceuticals Corp., P.O. Box 52330, Durham, NC 27711, USA		
TITLE	Location/Qualifiers		
FEATURES	1..1068		
SOURCE	/organism="Mus musculus"		
	/db_xref="taxon:10090"		
	/cell_line="Meth A"		
	/cell_type="fibrosarcoma"		
	1..1068		
	64..996		
mRNA	/standard_name="EMAP II"		
CDS	/codon_start=1		
	/product="endothelial-monocyte activating polypeptide II"		
	/db_xref="PID:9498912"		
	/translation="MATNDVLRRLDQGAEDQIIEYLKQVALIKERAILQATNKE EKLRVENAKLKEIEELKQVILAEIHNVQVRALSPLOTNCTASVSQSPV ATTAAPTRKQIRKAGEKKYKTEKKGEKKQGAASATDSKPIDASRLDRIGCI VTKKRRHDSILYVEVDVGEAAPRYVSLVNHVPEQONQNMVILCTLRKARG VLSQAMTCASSPEKVEIILAPNGSVPGDITITDAPGEDEKELNPKKIKWEDIQDL HTNAECVATYKGAFFEKVGGVCAQDMANSQIK"		
	496..993		
mat_peptide	/standard_name="EMAP II"		
	/product="endothelial-monocyte activating polypeptide II"		
polya_signal	1037..1042		
BASE COUNT	328 a 212 c 297 g 231 t		
ORIGIN			


```

/obj.xref="PID:92754814"
/translation="MCEYERLRPGTAPBSSPPAPATHRLRLGELRYRMRMTCRTV
AKATNDNVLNRLEQKGEADQIETELKQVALLKERVADLATREKKRLVENAKLKE
KEIELKQELQIAEIKONGVQKIPVQVSDPVOASNAVTSIVOSTSVTISCISKEB
SKGGEERKVEKPEKTDKPKKREKREKLOSLAPSDKRPVDSIDRIGRIVYKRPDPA
DLVEEEDVGEAAPRVYISGLVNHVPLDQONMRYVLCULKPKAKMGILSOAMVMO
ASSPEKEILLAPNGSVGDRIETFDAPFGEEDKELNPKKKIMEQIQDPLHNAECVAT
YKSGPEYKKGVCVCRADTMANSJK"
BASE COUNT      400 a      251 c      343 g      260 t
ORIGIN
Query Match      13.1%; Score 66; DB 14; Length 1254;
Best Local Similarity 64.1%; Pred. No. 1.15e-28;
Matches 150; Conservative 0; Mismatches 84; Indels 0; Gaps 0;
Db
681 TCTCGTCGTGATCTTCGATTTGGTCGTATTTCTACTGTCTCAAAAAGCACCCGGATCGACAG 740
109 TCCCGCGTGGATATCCGCTGTGGGAAAAATCATCATCTGTGGAGAACACCCAGATGCAGAC 168
741 TCACGTGATGTAGAGAAAGTGGATGTGGGGGAGGACGCCCAAGAGCAAGTCATCAGTGA 800
169 AGCGTGTATGTAGAGAAAGTATGACGTGGGGGAGGTGAACCAAGGACGTGTGTGAGCGGC 228
801 CTGGTGAATCATGTTCCCTCTCTAGACCAAGATGCAAAATCGAATGGTGGTATTCTTTGTAAT 860
229 CTGGTACAGTTCTGTCGCCCAAGAGGAACTGCAGCAGAGCGTGTGTGAGTGGCTGTGCAAC 288
861 CTGAAGCCCTGCAAAAATGGCGGATACTGTCTCAAGCAAGCTGATGTCCT 914
289 CTGAATCCCAAGATGAGAGAGTGCAGTCCCAAGGCGATCTTCTGTGTCT 342

```

BASE COUNT	577 a	318 c	335 g	473 t	TTTNDGLEVITFKDEERDHPVRKLTNNKGESEFKVASIANNQVR*
Query Match	6.5%;	Score 33;	DB 18;	Length 1703;	
Best Local Similarity	59.8%;	Pred. No. 9.53e-06;			
Matches	146;	Conservative	0;	Mismatches 95;	Indels 3;
					Gaps 1.
Db	1043	CCATCCGCAATAGATTTCGGTGTGTTTATTTCAAAAGCCATCAACACCCAGATGCC	1102		
Qy	106	CCATCCCGCGTGATATCCGTGTGGGAAAATCATCTGCGAGGAGACACCCAGATGCA	165		
Db	1103	GACGCCCATATGCTCTACAAATGATGTGGTATAGAAAGGCCCAAGACGTTGT	1162		
Qy	166	GACAGCCTGTATGTAGAGAGATGACGTGGGGAACTGAAA---CCACGCACTGTGTC	222		
Db	1163	TCTGTTTGTGCAAGCAATTTCTCTTGAGCCCTATGCAAGAACCTTATGTTGTGTGA	1222		
Qy	223	AGCGGCCCTGTACAGTTCGTGCCCAAGSAGSAACTGCAGAGCAGCGCTGATGTCTCTG	282		
Db	1233	TGCAACTTGAACACAGTTAACATGTAGAGGTTTAAATCCACTGCTATGTGTGTGTT	1282		
Qy	283	TGCAACCTGAAACCCACAGAAAGATAGAGAGAGTCCAGTCCCAAGCCATAGCTTGTGTGCT	342		
Db	1283	TCTA	1286		
Qy	343	TCTA	346		

[illegible]

RESULT	10	SCAR1	2473 bp	DNA	UNA	14-OCT-1996
LOCUS						
DEFINITION		S. cerevisiae	ARCI gene.			
ACCESSION		X95481				
NID		91620459				
KEYWORDS		ARCI gene.				
SOURCE		unidentified.				
ORGANISM		unclassified.				
REFERENCE		1 (bases 1 to 2473)				
AUTHORS		Simos,G., Segref,A., Fasiolo,F., Hellmuth,K., Sherchenko,A., Mann,M. and Hurt,E.C.				
TITLE		The yeast protein Arcp1 binds to tRNA and functions as a cofactor for the methionyl- and glutamyl-tRNA synthetases				
JOURNAL		EMBO J. 15 (19), 5437-5448 (1996)				
MEDLINE		97050848				
REFERENCE		2 (bases 1 to 2473)				
AUTHORS		Simos,G.				
TITLE		Direct Submission				
JOURNAL		Submitted (01-FEB-1996) G. Simos, University of Heidelberg, Institut fuer Biochemie I, Im Neuenheimer Feld 328, 69120 Heidelberg, FRG				
COMMENT		Related sequence: U31348.				
FEATURES		Location/Qualifiers				
source		1..2473				
		/organism="unidentified"				
		/db_xref="taxon:32644"				
		1..2473				
		/organism="Saccharomyces"				
		/strain="J04.2XJR26.19B"				
		/db_xref="taxon:4930"				
		554..1684				
		/gene="ARCI"				
		554..1684				
		/gene="ARCI"				
		/codon_start=1				
		/product="Arcp1"				
		/db_xref="PID:e221389"				
		/translation="MSDIYKFEELIISYVPSTKROSAQAQWESIVKSGOIQPHI				
		DQZLYLRDNTFIVSTIYPTSDVHYEVALPIKLDYASSKDYKSTYTRRLRLMT				
		DYQNLVLEVSSTDKLEINHDDLPEHYIEKKKAPAGGADPAAKADSDYKKKKQD				
		HPKRPDEETLTKLREBAKAKKAAKAAQAQOQNKAAPEKRPALDFVGFQK				
		AIKRPDSDIYVSTIDVGDEGPTVCSGLVHKPFLDMQERYVVVVCNLKPVNMRGI				
gene						
CDS						

[illegible]

Db	10829	10889	10929	10934	10938	10944	10948	10954	10958	10964	10968	10974	10978	10984	10988	10994	10998	11004	11008	11014	11018	11024	11028	11034	11038	11044	11048	11054	11058	11064	11068	11074	11078	11084	11088	11094	11098	11104	11108	11114	11118	11124	11128	11134	11138	11144	11148	11154	11158	11164	11168	11174	11178	11184	11188	11194	11198	11204	11208	11214	11218	11224	11228	11234	11238	11244	11248	11254	11258	11264	11268	11274	11278	11284	11288	11294	11298	11304	11308	11314	11318	11324	11328	11334	11338	11344	11348	11354	11358	11364	11368	11374	11378	11384	11388	11394	11398	11404	11408	11414	11418	11424	11428	11434	11438	11444	11448	11454	11458	11464	11468	11474	11478	11484	11488	11494	11498	11504	11508	11514	11518	11524	11528	11534	11538	11544	11548	11554	11558	11564	11568	11574	11578	11584	11588	11594	11598	11604	11608	11614	11618	11624	11628	11634	11638	11644	11648	11654	11658	11664	11668	11674	11678	11684	11688	11694	11698	11704	11708	11714	11718	11724	11728	11734	11738	11744	11748	11754	11758	11764	11768	11774	11778	11784	11788	11794	11798	11804	11808	11814	11818	11824	11828	11834	11838	11844	11848	11854	11858	11864	11868	11874	11878	11884	11888	11894	11898	11904	11908	11914	11918	11924	11928	11934	11938	11944	11948	11954	11958	11964	11968	11974	11978	11984	11988	11994	11998	12004	12008	12014	12018	12024	12028	12034	12038	12044	12048	12054	12058	12064	12068	12074	12078	12084	12088	12094	12098	12104	12108	12114	12118	12124	12128	12134	12138	12144	12148	12154	12158	12164	12168	12174	12178	12184	12188	12194	12198	12204	12208	12214	12218	12224	12228	12234	12238	12244	12248	12254	12258	12264	12268	12274	12278	12284	12288	12294	12298	12304	12308	12314	12318	12324	12328	12334	12338	12344	12348	12354	12358	12364	12368	12374	12378	12384	12388	12394	12398	12404	12408	12414	12418	12424	12428	12434	12438	12444	12448	12454	12458	12464	12468	12474	12478	12484	12488	12494	12498	12504	12508	12514	12518	12524	12528	12534	12538	12544	12548	12554	12558	12564	12568	12574	12578	12584	12588	12594	12598	12604	12608	12614	12618	12624	12628	12634	12638	12644	12648	12654	12658	12664	12668	12674	12678	12684	12688	12694	12698	12704	12708	12714	12718	12724	12728	12734	12738	12744	12748	12754	12758	12764	12768	12774	12778	12784	12788	12794	12798	12804	12808	12814	12818	12824	12828	12834	12838	12844	12848	12854	12858	12864	12868	12874	12878	12884	12888	12894	12898	12904	12908	12914	12918	12924	12928	12934	12938	12944	12948	12954</
----	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	---------

Qy 328 ATGCTTGTGCTGCTCA 346

RESULT 13 128278 215 bp DNA PAT 30-OCT-1996
DEFINITION Sequence 5 from patent US 5569830.
ACCESSION 128278
MID g1819054
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS Bennett, A., Labavitch, J.M., Powell, A. and Storz, H.
TITLE Plant inhibitors of fungal polygalacturonases and their use to control fungal disease
JOURNAL Patent: US 5569830-A 5 29-OCT-1996;
FEATURES
source Location/Qualifiers
1. 215
BASE COUNT 15 a 8 c 25 g 26 t 141 others
ORIGIN

Query Match 6.2%; Score 31; DB 25; Length 215;
Best Local Similarity 14.3%; Pred. No. 1.59e-04;
Matches 22; Conservative 64; Mismatches 67; Indels 1; Gaps 1;

Db 1 MNNVMSSSVYRSTRSCNDKAKKNDQNTSSWTIDCCNRTMVCOTDITTYRNNDGSHK 60
Qy 395 GTGCTCGGTGACGACGCTGTTGTGAAGGCTATGAAAGGCGCAACCGATGAGAGC 454
Db 61 YSSANYTGNVNGAATHTYHTNVSGADSKVTDSYNASGSSNGTGDNRSGADSY 120
Qy 455 TCAAGCCCAAGAGAAGTCTTCGAGAGTTCAGAGCTCAAAATTTCTGAGAGCT 514
Db 121 GSSKTAMTSRNTGKTANNANDSRNMGASVGS 154
Qy 515 GCATCGCACA-GTGGAGCAACCAACTCATGA 547

RESULT 14
LOCUS ECHE000389 11157 bp DNA 02-SEP-1997
DEFINITION Escherichia coli K-12 MG1655 section 279 of 400 of the complete genome.
ACCESSION AE000389 U00096
MID g1789451
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS Blatter, F.R., Plunkett III, G., Bloch, C.A., Perna, N.T., Burland, Y., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B. and Shao, Y.
TITLE The complete genome sequence of Escherichia coli K-12
JOURNAL Science 277 (5331), 1453-1474 (1997)
MEDLINE 97426617
REFERENCE
AUTHORS Blatter, F.R.
TITLE Direct Submission
JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
3 (bases 1 to 11157)
Blatter, F.R.
Direct Submission
Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459

COMMENT

The E. coli K-12 sequence and its annotations have been updated. All of the ambiguous residues in our original submission have been resolved, and mis-assemblies in two repetitive regions have been realigned. The annotations have been improved and updated as well. With this release we begin designating a version number for the annotated sequence, to assist in keeping track of corrections, updates, and other changes. This is version M52 (SEP. 02, 1997). In addition, a revised notation has been instituted which assigns each gene (protein- or RNA-encoding) a unique numeric identifier beginning with a lowercase 'p' (in the '/label' field); this will remain constant through further updates, gene identifications, etc. This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHRR). The entire sequence was independently determined from E. coli K-12 strain MG1655. Predicted open reading frames were determined using Genemark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332. e-mail: mark@amber.gatech.edu

Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (<http://cgsc.biology.yale.edu>). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (<http://www.genetics.wisc.edu>).

FEATURES

source

1. 11157
/organism="Escherichia coli"
/strain="K-12"
/sub_strain="MG1655"
/db_xref="taxon:562"
174. 797
/note="0207: This 207 aa ORF is 42 pct identical (7 gaps) to 52 residues of an approx. 200 aa protein SLTD_ECOLI SW: P30856"
/codon_start=1
/label-b3071
/db_xref="PID:g1789452"
/transl_table=11
/translation="MSHHHEGCCCHHEGPPHHEGCCKEKSEHNECGHGHHEGCCG GRHGGGGRQRFEGHGLRVLDILDSRDSHSEYLKAIENLTQGNVPPSPVITP TLDFLOBSLITFREEGKQIALTFEGAGWLENEDEOYEITEIRAKVCAGALRO NPOKRALDNFKAVLDLRVNSDSDAQIKKIIAVIDRAAFDITQLD"
complement(951. .2471)
/gene="air"
complement(951. .2471)
/gene="air"
/note="f506; 100 pct identical to AIR_ECOLI SW: P50466"
/codon_start=1
/label-b3072
/product="aerotaxis receptor protein"
/db_xref="PID:g1789453"
/transl_table=11
/translation="MSSHPYVTOQNTPLADDTLMSTTDQSYTTHANDTFVOYSGYT LOELGQPHNNVRHPDMKAFADMMFTLKGEPSGIVKRRKNGHWYRAAVPM VREKISGYSIRPRADDEIAVEPIYKRLNGRSTRKHGKIVYKGLGRIPSLP LRRARQVNTLMPTLLAAMPFAAPVYVITILCAUUVLLASACENQIYRPIEVAAO ALKVAITERNSVEHLNRSDELGLTRAVGUGLGMVLINDVSSQVSSVNGSSTLAK GTBELNHTQOTVDNVQOYVATNMQMAVSKNSATNSADKSLTSSNAVOGGMK TTVIKTMDIDTADSFORTGTTSLINDINQALALAAVEAARAGOGGCFVAVAE VRHLASANAANDIRKLIDASADKVSQGVAAARTMEDIVAAQVKNTQTLAQS HSTLEQADGLSSLRAVDELNIITQKNAELVEESAQVSAVYKRRASLDEAVYVTLH" 2778. 4268
/gene="y9jG"
2778. 4268
/gene="y9jG"
/EC_number="2.6.1.13"
/note="0496; 100 pct identical to 429 amino acids of


```
/sub_strain="MG1655"  
/note="This sequence comprises the following lambda  
clones: Dd693(EC30K511-4), Dd695(EC27-779)  
Dd698(EC30K513A-10pp), Dd701(EC30K514A-6pp),  
Dd703(EC30K514A-4pp), Dd704(EC27-355), Dd707(EC30K518-4),  
Dd709(EC30K518-5), Dd712(EC17-155), Dd713(EC30K520-13pp),  
Dd715(EC30K520-15pp), Dd718(EC27-977), Dd721(EC30K523-9),  
Dd723(EC30K524-3pp), Dd725(EC27-1316), Dd727(EC30Mm23),  
Dd730(EC27-417), Dd731(EC18-29), Dd734(EC19-161),  
Dd737(EC5-106), Dd741(EC30WRA4), Dd743(EC27-1299),  
Dd744(EC29-14), Dd747(EC22-12), Dd750(EC17-214),  
Dd751(EC18-285), Dd753(EC17-149), Dd757(EC14-35),  
Dd758(EC17-5), Dd760(EC27-516), Dd762(EC30K519-15pp),  
Dd763(EC30Mx35), Dd765(EC30Mx31), Dd766(EC27-394), and  
Dd770(EC18-109); M3mplr or Janus vectors were used for  
subcloning".  
/db_xref="taxon:562"  
/map="67.4 to 76.0 minutes"  
1..112  
/note="Predicted bend of 75.70 degrees"  
136..1626  
/note="ORF .q496"  
/codon_start=1  
/db_xref="pid:g606011"  
/transl_table=11  
translation="MTTEFEPIPAAGVQSLQLYLNRYHVRREPHILNRLPSS  
ASALCSAHMLLEKRLLIDHEEMALREVIEYEKEHNPGFLERKSYPAGSDYA  
VEWGQSNTLVDPGGOEFDLGGGFIFNVGHRNPVVVASOVOLKOPPLHSOELD  
PRWLATLTIALAVPTGRKTSFGPCSGSESVREALKLAKAVSRGRFTPLTSGAFR  
CKSLGLSATAKSTFRKPFMPNLBGRFPVGNGEAMRTALECKTKGDVAAYILEP  
IQEGGVLPPEGYITAAVKLDDEFGMALLDEVOTQGRRGKKACAEHEHVQDILLC  
LAKAAGGVMPGATIGALDEEVSYLFDMPLFTTFGNGPLCAKAALAATINVLQONI  
PQAOKRCMDMLDGROLAREPDLYODARKGMMALEFYVDNEIGNVFSEMFRQVR  
LVAGLNNAKTRIIRPLETLTIEQEELVIKARALAMARYSBEA"  
complement(1668..2000)  
/note="ORF .f110"  
/codon_start=1  
/db_xref="pid:g606012"  
/transl_table=11  
translation="METVAVADPARLEMVGKIYEVKHENADKITIVQDVYGKTLO  
TVTSIIVPYSEEELMKTIYYVLCLNRKKNMGFTSECHLCLETDDGESVELTPERRK  
MPAGRKYV"  
2094..7072  
/note="corresponds to GenBank Accession Number M64441,  
ECOBERGRA"  
2094..7072  
/note="corresponds to GenBank Accession Number X52031,  
ECEBG"  
2219..6461  
/note="corresponds to GenBank Accession Number X03228,  
ECBERGRA"  
2219..3202  
/gene="ebgr"  
2219..3202  
/gene="ebgr"  
/note="CG Site No. 829"  
/codon_start=1  
/function="ebg repressor"  
/db_xref="pid:g606013"  
/transl_table=11  
translation="MATLKDAIEAGVSIAIVSRYLNDPTLVNKETKRHLIEIAEK  
LEYKTSRKRLTGAVNOHHILAISYOOLEINDPYLIARHGIETCEKLGIELNNK  
CYEHSGEDIKVTGTGIIIVGKRPALPRAASALEDNICFIDEHPGSGDYVDIDLAR  
ISKELIIDYINGGNVRIIGFIGEDEPGCAADI REVAFAHYGKLOYRREDIWRGFS  
SSGTGLAKOMLRDEDYPRALFYASDSINAIGVARLHEGVLNIPODISLISVNDIPLAR  
FFEPPLSTAIRIHSEMWGSQGVNLVERKADERALRVLEVSKLRLCTRP"  
3383..3383  
/note="predicted bend of 87.49 degrees"  
3350..6478  
/gene="ebga"  
3350..6478  
/gene="ebga"  
/note="CG site no. 830"
```

/codon_start=1
 /product="phospho-beta-D-galactosidase; alpha-subunit"
 /db_xref="PID:g606014"
 /transl_table=11
 /translation="MNGSDSESLPGGPMNRNENIQLEHNRLAPAYPFESYSVAOATP
 FARENSSTFELSGOAWNHPFDPHQLQPEATSTSLMDKGIITVPRAMOWGHRKLOY
 TDEGPEPPIIDPEVPSDNPICATIORITTLSDGQKOTTLIKFDVSEITYEYYNGOIV
 GFSGKSRITLAFDISAWKTDGDLICVRVMQADSTVEEDDMQVETIFRDYLVK
 HLTHINDTVAFTEDEAVACDATALCEVLEMLASPVYTLLEYLFGGERVHSSAID
 HLALEKLSAFAFTEVDEOPOOMASESPILYHLVMTLIDANENYLEVYQORGFEDIVY
 RDGLFVNNRWYHLGVRHNDHRKGRGAVMDRVEDLOLMKQHNINSTRAYLPND
 RPVELCDIYELFLYMAETDVSHCFANVGDISRTDDPQWKEVYEVYVIRIHAQRKH
 PSIIWMSGNSGVCNTRMTAAKALDTRLVTHIEDRDAEYVDIISTRTKPLML
 PSIIWMSGNSGVCNTRMTAAKALDTRLVTHIEDRDAEYVDIISTRTKPLML
 NGEVYFPPKRRICCEYAHAMNGPBGGLTEQONVYKHCIOGYWECMDHGLQADQ
 DHGWAYFPGGDCYDPPNNYFNCGLDITVSPQTPGLKEKYIAVKAJHARLQTEG
 ELKENVKSTLTDYTLLHAVERAEGELTAQIKLDVAPNSAPQITLPOLADESE
 AFINETVYKDSRTTSEAGHPILATYQFPIKENTQOPVPAFNNARPLIEDRSCVIV
 RGNVATIRFSKMSKSPKTSWYNGESLITREKINFPKPHMDNHQVEGLMQPHLOID
 MOEHLRDAVAFQSGEVLILSRVYIAPPEVPGKCYITMILADGVAWVALSERG
 DYPHIPIGTFQMGINEGYDQVAYGRGPENYADSOANITIDRQAVPMANFENYF
 PONGNRRWETALTNHRHNGGLVYVQRPINFEAMHYIOTENIIAACHQELQSDSDI
 TLNDHOLLGAGTSMGSEVLDISRWVRDPSYGTLLPVSGEATQASOLASYEFBGA
 FESNLSHENQ"
 6475..6924
 /gene="ebgc"
 6475..6924
 /gene="abgc"
 /note="CG Site No. 18412; difference near end makes this
 25 amino acids shorter than previously published
 sequences"
 /codon_start=1
 /product="phospho-beta-D-galactosidase, beta-subunit"
 /db_xref="PID:g606015"
 /transl_table=11
 /translation="WRIDNLEQFROIYASKKQRQVEALEINDINQPGVHSGTGS
 LTRVEDASDALEFTGHRRYFEVHYILOGQKLEIAPKETLOVEYVIRDETREYLK
 GCGEIVHEQOQIYICDIEHAYRFICNNAYKRVYLVKTIEDGYPHNK"
 6987..8420
 /note="ORF 0477"
 /codon_start=1
 /db_xref="PID:g606016"
 /transl_table=11
 /translation="MSDKRNTIGKFGILLSTPFAVYSFNNVINNNIELGASAPMEF
 LATIYEPFCLILIAEFVSLNKNSDAQYAVKSLSGRMATFAYTYEWNLEFFSL
 LLPRIVASAFIAGEYEXYIMPTVATITISMLFAPNIVYNSNGMKGLPITSVTLML
 LLTSLTILAGTALNGYVOPADAITDAMIPNNMAYPLCTYTMIAAGASVAYY
 VNDVSGKSTCVKYLILAGITGVLYSVSYLVNVEYSKRLATGTGSSOVYFHMAIY
 FGLPEALNVRGVLVSFTAMFGSLMTATPVKLFEEIPGIEGKTATGSSOVYFHMAIY
 RSLVDVNRDRPADYDPMLGSNTOVDLMTNITINNTAASMLPPLFIMLAYINLEAKLID
 HLPREFRGSRRGTGIIVSMILAEFAGVASTPTGANNILTIIFYNVGIVIFLGP
 MMKYSKYIKGLAEERHLEATPSANVD"
 8554..9624
 /note="ORF 0356"
 /codon_start=1
 /db_xref="PID:g606017"
 /transl_table=11
 /translation="MKLTAPCALIALPFCYARFASAGEEARPAEHDTKPTATTS"
 SPSPRFYSELVCGMKMLEGKNKHKYSNGYITEGLEMKYSMPGLYIGBQYVADQAH
 DGNANVPYHSHWGGEFGGINRRYGGIGRIDGTEIMLSLRDSSLDLDMQWDFITDLOT
 YIPNTRDITMLAKVQLNSGNRFYSVTAIPAHHDSEKAWMLFHGTGYDRDYDYTPPAM
 NGYIOLDABEITVMNGLEITDGTGOLITGLIPNTPAARMAHHTGRADGIDYVYSS
 GMAVAMAEALKGVLISATYAYAKHRHPHADDETSFMOGFIWYEGGGRATAFDRSR
 FYMKASHPDSDQIFLMQYFYW"
 9641..11992
 /note="ORF 0783"
 /codon_start=1
 /db_xref="PID:g606018"
 /transl_table=11
 /translation="MKRITLITPVTCALLISFSAHANADNYKAVINRTGAPQVYMKDY
 DYDQQRNPNFPDGGAHGHLIPDQGVNTMGFPYVALITREYINFAFNSDRPLWKKY
 GKVDVTEA'SIPALVQKLTAKDOVQEMTLRRATPSTLLEKTI'SNKPFLDIWVG
 ELKELEKEKPSLSDKTIAGEYDPYQKTSATDGLKVFYGGKRAVWATLDTGSESESY
 QVHKSIPQVTEINPGRSTKINSINGSTVLYVYSHLLTAQGVSEKQOIRDIILAPPEF

ORGANISM	Homo sapiens Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Euthera; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 525)
REFERENCE	Hillier, J., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Persons, J., Rifkin, L., Rohlfing, T., Tan, F., Treaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
TITLE	Mashu-Merck EST Project
JOURNAL	Unpublished (1995)
COMMENT	Contact: Wilson RK Mashu-Merck EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available royalty-free through LMNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Seq primer: -28M13 rev2 from Amersham High quality sequence stop: 330. Location/Qualifiers 1..525
FEATURES	/organism="Homo sapiens" /contig="Organ: colon; Vector: pBluescript SK-; Site.1: ECORI; Site.2: XhoI; Cloned unidirectionally. Primer: Oligo dt. T-84 colonic epithelial cell line. Average Insert size: 1.0 kb; UniZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCACAG 3' -3' adaptor sequence: 5' CTGACGTCTTTTGTATTTTTTTT 3'. /db_xref="taxon:9606" /_clone="511328" /_clone_1lb="Stratagene colon (#937204)" /_lab_host="SOBR cells (kanamycin resistant)" <1..>525
BASE COUNT	mRNA ORIGIN 144 a 125 c 149 g 105 t 2 others
Query Match	60.8%; Score 363; DB 15; Length 525;
Best Local Similarity	96.5%; Pred. No. 0.00e+00;
Matches 412; Conservative	0; Mismatches 8; Indels 7; Gaps 7;
Db	62 TACCCTGCCTCGAAAAAATGCGCCAGCGTGCTTACCAGATCCCTCAAGCAGAAGCC 121
Oy	1 TACCCCTGCCTCGAAAAAATGCGCCAGCGTGCTTACCAGATCCCTCAAGCAGAAGCC 60
Db	122 AATGCCAAACCATGCGCAAAATTGCAAGAACCAAGAGAGATCATCCCATCCGGCTGGAT 181
Oy	61 AATGCCCAAAGCGCTGCCAAGAAATTCAGAAACAAGAGAGATCATCCCATCCGGCTGGAT 120
Db	182 ATCCGTGTGGGGAATAATCATCTACTGTGGAGAGACCAAGTCAGACAGACGCTGTATGTA 241
Oy	121 ATCCGTGTGGGGAATAATCATCTACTGTGGAGAGACCAAGTCAGACAGACGCTGTATGTA 180
Db	242 GAGAAGATTGACGTGGGGGAAAAGCTGAAACACAGACTGTGTTGTTGAAGCGGCTGTACAGTT 301
Oy	181 GAGAAGATTGACGTGGGGGAAAAGCTGAAACACAGACTGTGTTGTTGAAGCGGCTGTACAGTT 239
Db	302 CGTCCCCAAGAGAATAATTCAGACGACAGGCTGTGTGTGTGTGTGCAAACTGAAACCCCA 361
Oy	240 CGTCCCCAAGAGAATAATTCAGACGACAGGCTGTGTGTGTGTGTGCAAACTGAAACCCCA 299
Db	362 GAAGATGAGAGGAGTAGAGTCCCAAGGCATGCTTGTGTGTGTGTCTTATAGAAAGGGATA 421
Oy	300 GAAGATGAGAGG-AGTCGAGTCCCAAGGCATGCTTGTGTGTGTGTCTTATAGAA-GGGATA 357
Db	422 AACGCCAGATTGAACCTCTGTGAACTCTCCGGCAAGGCTCTGCTCCCTGGTTAGACAGT 481
Oy	358 AACGCCAGGTTGAACCTCT-TGACCTCTCGGCA-GGCTGTGCTCC-TGGT-GAGCAGCT 413
Db	482 TGTTTG 488

Oy	414	GTTTGTG	420
RESULT	4	R18984	407 bp mRNA EST 14-APR-1995
LOCUS		yq25d10.r1 Homo sapiens cDNA clone 33689 5' similar to SP:SYM_THERRH	
DEFINITION		P23395 METHIONYL-TRNA SYNTHETASE ;.	
ACCESSION		R18984	
NID		g772594	
KEYWORDS		EST.	
SOURCE		human clone-33689 library-Soares infant brain lNIB vector-Lafmid BA host-H10B (ampicillin resistant) primer-M13R1 RstEcl-Not I RstE2-Hind III whole brain from a 73 days post natal female. 1st strand cDNA was primed with a Not I - Oligo(dT) primer [5'] AACTGGAAACAATCCGGCGCAGCAATTCTTTTTTTTTTTT 3'; double-stranded cDNA was ligated to Hind III adaptors (Pharmacia), digested with Not I and directionally cloned into the Not I and Hind III sites of the Lafmid BA vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo.	
ORGANISM		Homo sapiens Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;	
REFERENCE		Eutheria; Primates; Catarrhini; Homnidae; Homo.	
AUTHORS		1 (bases 1 to 407) Hallier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucabz,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.	
TITLE		The Mashu-Merck EST Project	
COMMENT		Unpublished (1995)	
		GDB: G00-406-036 Contact: Wilson RK Mashu-Merck EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel.: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu High quality sequence stops: 294 Source: IMAGE Consortium, LNLN This clone is available royalty-free through LNLN ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Location/Qualifiers 1..407 /organism="Homo sapiens" /clone="33689"	
FEATURES		source	
BASE COUNT		103 a 87 c 121 g 95 t 1 others	
ORIGIN			
		Query Match 48.4%; Score 289; DB 10; Length 407; Best Local Similarity 94.5%; Pred. No. 0.00e+00; Matches 364; Conservative 0; Mismatches 10; Indels 11; Gaps 11;	
Db	1	AAGATTGACGTGGGGGAAGCTGAACACAGCATGTGTGAGCGGCGCTGGTACAGTTCTG 60	
Oy	184	AAGATTGCAGTGGGGGAAGCTGAACACAGCATGTGTGAGCGGCGCTGGTACAGTTCTG 243	
Db	61	CCCAAGAAGGAACATGCGAGACAGGCTGTAGTGTGCTGTGTCGAACCTGAAAACCCAGAA 120	
Oy	244	CCCAAGGAGGAATGCGAGACAGGCTGTAGTGTGCTGTGTCGAACCTGAAAACCCAGAA 303	
Db	121	ATGAGAGAGTCGAGTCCCAAGGCGATGCTTGTGTGCTTCTATAGAAAGGATAAACCGC 180	
Oy	304	ATGAGAGAGTCGAGTCCCAAGGCGATGCTTGTGTGCTTCTATAGAAAGGATAAACCGC 363	
Db	181	CAGTTGANACCTCTGGAACCTCCGGCAGAGCTTCTGCTTCCCTGGGTGAGCACAGTGTGTG 240	
Oy	364	CAGGTGTAACCTCTGGAACCTCCGGCAGAGCTTCTGCTTCCCTGGGTGAGCACAGTGTGTG 420	
Db	241	AAGGCTATGAAAAGGGCCAACGATGAGAGAGCTCAAGCCCAAGAAGAAAGCTTTTTCGA 300	

QY	421	AAGGCTATGAAAAAGGGGCAACCGATGAGGAGCTCAAGCCCAAGAAGAAAGCTTT-CGA	479
Db	301	GAAATTTGCAGAGCTTGACTTTCAAAATTTCTTAGAGAGTGTCATTCGCAAGTTGGGAAGCA	360
OY	480	GAAATTT-GCAGGGCT-GACTT-CAAAATTTCTT-GAGGAGTGTCAT-CGCACAGCTGGG-AAGCA	533
Db	361	AACCACTTTCATTCAGCCAAGTTGGG	385
OY	534	AACCACTTTCAT-GACCAAGCTGGG	557
RESULT	5		
LOCUS	AA102052	459 bp	mRNA
DEFINITION	2178412.r1 Stragelene colon (#937204) Homo sapiens cDNA clone		
ACCESSION	AA102052		
NID	91645892		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 459)		
AUTHORS	Hillier,L., Clark,N., Dubuque,T., Elliston,R., Hawins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifein,L., Rohlfing,T., Tan,F., Treaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.		
TITLE	Washn-Merck EST Project		
JOURNAL	unpublished (1995)		
COMMENT			

Contact: Wilson RK
Washu-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
This clone is available royalty-free through LINL ; contact the
IMAG Consortium (info@image.linnl.gov) for further information.
Seq primer: -28kb1 rev2 from Amersham
High quality sequence stop: 418.

FEATURES	Location/Qualifiers
source	1. .459

```

/organism="Homo sapiens"
/node?Organ: colon; Vector: pBluescript SK-; Site.1:
EcoRI; Site.2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. T-84 colonic epithelial cell line. Average
Insert size: 1.0 kb; Uni-ZAP XR Vector; ~5' adaptor
sequence: 5' GAATTCGCACACG 3' ~3' adaptor sequence: 5
CTCGAGTTTTTTTTTTTTT 3'
/db_xref="taxon:9606"
/clone="510790"
/clone.lib="Stratagene colon (937204)"
/lab_host="SOLR cells (kanamycin resistant)"
<1..7453
mrna
BASE COUNT
114 a 118 c 117 g 107 t 3 others

```

Query Match	45.6%	Score 278;	DB 15;	Length 459;
Best Local Similarity	95.6%;	Pred. No. 0.00e+00;		
Matches	368;	Conservative	0;	Mismatches 2;
			Indels 15;	Gaps 15

Db	1	AGCGCCCTGGTACAGTTCGTGCTCCCAAGAGAACTGCAGGACAGCGCTGTAAGTGCTCG	60
QY	223	AGCGCCCTGGTACAGTTCGTGCTCCCAAGAGAACTGCAGGACAGCGCTGTAAGTGCTCG	282
Db	61	TGCACCTGAAACCCCGAAGATGAGAGAGTGAATGCCAAGGCATGCTTGCTGCT	120
QY	283	TGCACCTGAAACCCCGAAGATGAGAGAGTGAATGCCAAGGCATGCTTGCTTGCT	342
Db	121	TCATATGAAGGATAAACCGCGAG-TTGAACCTCTGACCTCCGCGAG-CTCTGCTCT	178

QY	343	TCCTATAGAGAGGATATAAACCCGACAGTTGAACCTCTGGACCCCTCCGAGGCTCTGCTCT	402
Db	179	GGTAGCACCCTGTTTGTGAAAGGGCTATATGAAAGGCCAACAGATATAGAGCTCAAGCC	238
QY	403	GGTAGCACC-GTGTTTGTGAAAGGGCTATGAAAGGCCAACAGATATAGAGCTCAAGCC	461
Db	239	CAAGAGAAAGTCTTCGAGAAAGTTCGAGGCTGACTTCAAAATTTCTTGAAGAGATCAT	298
QY	462	CAAGAGAAAGTCTTCGAGAA- GTTCGAGGCTGACTTCAAAATTTCTT- GA- GGAGTGCAT	518
Db	239	CNCACAAGTTGGAAACCAACCACTTTCATTTGACCAAGCTTGGGCTCCATTTCCCTGTGA	358
QY	519	CGCACA- GT- GGAAGCAACCAACTT- CAT- GACCAAGCT- GGGCTCCATTTCC- TGT- A	571
Db	359	AATGCGTTGAAAGGGGAGCAATT	383
QY	572	AATGCGTT- GAAA- GGGGGGAACATT	594

LOCUS	6	AA221081	517 bp	MRNA	EST	12-FEB-1997
DEFINITION		smv65f03.r1 Soares mouse 3MME12 5 Mus musculus cDNA clone 659933 5				
ACCESSION		AA221081		GI184659	TYROSYL-TRNA SYNTHETASE. ;	
NID		G1840276				
KEYWORDS		EST.				
SOURCE		house mouse.				
ORGANISM		Mus musculus				

REFERENCE	AUTHORS	TITLE	COMMENT
1 (bases 1 to 517)	Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubouque, T., Gelsel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Weising, B., Wylie, T., Lennon, G., Soares, B., Watson, R. and Waterston, R.	The WashU-HHMI Mouse EST Project	Unpublished (1996)

Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@lml.gov) for further information.
MGI:405781
Seq. primer: -28m13 rev2. ET from Amersham
High quality sequence stop: 453.

FEATURES	Location/Qualifiers
source	1. .517

```

/organism="Mus musculus"
/strain="C57BL/6J"
/notes="Vector: p7r13D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTCCAAATCTGAAGTGGAGGCGCCGCTTATTTTTTTTTTTT 3'],
on total mRNA provided by Minoru Ko, Wayne State
Univ.;] double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified p7r13 vector.
Library went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
/db_xref="taxon:10090"
/clone="659933"
/clone_1lb="Soares mouse 3NME12 5"
/sex="unknown"
/dev_stage="12.5dpc total fetus"
/lab_host="DH10B"
<1. -517

```

miRNA


```

source
1. 243
/organism="Homo sapiens"
/note="vector: lambda ZAP Express; Site_1: EcoRI. Site_2:
XhoI. mRNA was purified from human fetal hearts (8-10
weeks). cDNA was synthesized using a XhoI-01190 dT
adapter-primer. EcoRI adaptors were ligated, followed by
digestion with XhoI, for directional cloning into
predigested lambda ZAP Express."
/db_xref="taxon:9606"
/clone="L1914"
/lab_host="E. coli XL1-Blue"
BASE COUNT 57 a 58 c 77 g 51 t
ORIGIN
mRNA
Query Match 36.7%; Score 219; DB 14; Length 243;
Best Local Similarity 96.7%; Pred. No. 0.00e+00;
Matches 237; Conservative 0; Mismatches 6; Indels 2; Gaps 2:

Db 1 CCCAGATGCAGACAGCCTGTATGTAGAGAAGATTGACGTGGGGAGCTGAACCGGAC 60
|||
|||
|||
Oy 156 CCCAGATGCAGACAGCCTGTATGTAGAGAAGATTGACGTGGGGAGCTGAACCGGAC 215
|||
|||
|||
Db 61 TGTGGGACGGCGCTGTACAGTTCGTGCCCAAGGAGAACTCGAGACAGCGCTGTAGT 120
|||
|||
|||
Oy 216 TGTGGGACGGCGCTGTACAGTTCGTGCCCAAGGAGAACTCGAGACAGCGCTGTAGT 275
|||
|||
|||
Db 121 TGTGCTGTTCACACTGAACCCGAGAAATGAGAGAGTGCAGTCC-AAGCATGCTTCT 179
|||
|||
|||
Oy 276 GGTGCTGTTCACACTGAACCCGAGAAATGAGAGAGTGCAGTCCCAAGCATGCTTCT 335
|||
|||
|||
Db 180 GTGTCTTCTATAGAAGGATAA-CCGCCAGGTTGAACCTCTTGGCCCTCCGGCAGGTT 238
|||
|||
|||
Oy 336 GTGTCTTCTATAGAAGGATAAACCGCCAGGTTGAACCTCTGACCCCTCCGCGAGGCTC 395
|||
|||
|||
Db 239 TGCTC 243
||||
Oy 396 TGCTC 400
||||

RESULT 9
LOCUS AA003398 423 bp mRNA EST 19-JUL-1996
DEFINITION mg57b09.r1 Soares mouse embryo NBME13.5 14.5 Mus musculus CDNA
clone 437081 5' similar to PIR.A55053 A55053
endotherm11homonocyte-activating protein II precursor - mouse ;.
ACCESSION AA003398
NID g1446863
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Euteleostomi; Eumetazoa; Chordata;
Vertebrata; Lutherochordata; Rodentia; Sciurognathi; Myomorpha; Muridae;
Murinae; Mus
1 (bases 1 to 423)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisler,S., Kucaba,T., Lacey,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HM Mouse EST Project
Unpublished (1996)

TITLE
JOURNAL
COMMENT
Contact: Maria M/Mouse EST Project
WashU-HM Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.wustl.edu
This clone is available royalty-free through LML; contact the
IMAGE Consortium (info@image.lml.gov) for further information.
MGI:262417

```

Seq primer: EYPrimer
High quality sequence stop: 343.
Location/Dualifiers

1. .423 /organism="Mus musculus"
/note="Vector: pT7R3D-Pac (Pharmacia) with a modified polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - Oligo(dt) primer [5', TGTTACCAATCTGAAGTGAGCGCCGCCGAAATTATTTTTTTTTTTTTTTT T 3'] , on equal amounts of mRNA from 2 13.5dpc and 2 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne State Univ., from 2] ; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested wih Not I and cloned into the Not I and Eco RI sites of the modified pRTV3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."
/db_xref="taxon:10090"
/clone_id="Soares mouse embryo NbME13.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
<1. .>423

mRNA BASE COUNT 98 a 115 c 120 g 90 t

ORIGIN

Query Match Best Local Similarity 36.0%; Score 215; DB 14; Length 423;
Matches 241; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

Db 157 GAACTCGAGCAGGAGCGCTGTGTGCTGCATCATCTGAATAAACCAGAGATGAGAGC 216
|||||
Oy 253 GAACTCGAGCAGCAGGAGCGCTGTGTGCTGCATCATCTGAATAAACCAGAGATGAGAGA 312
|||||

Db 217 GTGACTCGCAGGGCATCTACTGTGCTCTTGTAAGAGGGGTGAGCCGCCAGTTAA 276
||| |
Oy 313 GTCGATGCCAAGGAGTCTCTGTGTCCTTCTATAGAAGGAGATPAAACCCGAGGTTAA 372
||| |

Db 277 CCTGTGACCCCTCCGCGCTGCGTCCGTCGAGTAGAGAGTTCGTACAGGGCATGAG 336
||| |
Oy 373 CCTGTGACCCCTCCGCGCTGCGTCCGTCGAGTAGAGAGTTCGTACAGGGCATGAG 432
||| |

Db 337 AAGGCCAGCCAGATGAGAGACTCAAGCCCAAAGAAAGCTTTGAGAGACTGCAGGCC 396
|||||
Oy 433 AAGGCCAACCAAGATGAGAGACTCAAGCCCAAAGAAAGCTTGCAGAACTTGAGAGCT 492
|||||

Db 397 GAC TTAAA ATT TC TG AGA GT GC AT C 423
|||||
Oy 493 GAC TTC AAA ATT TC TG AGA GT GC AT C 519

RESULT 10 LOCUS H13315 437 bp mRNA EST 27-JUN-1995
DEFINITION y172b07.r1 Homo sapiens cDNA clone 43366 5' similar to
SP:SYSC_YEAST P36421 TYROSYL-T-RNA SYNTHETASE, CYTOPLASMIC ; .
H13315
NID 9878135
KEYWORDS EST.
SOURCE human clone=43366 library=Soares infant brain JNTB vector-Lafmid BA
host=DHI0B (ampicillin resistant) primer-M13RP1 Rsite1-Not I
Rsited-Hind III Whole brain from a 73 days post natal female. 1st
strand cDNA was primed with a Not I - Oligo(dt) primer [5'
AACTGGAAAGATTCGGCGCCGAGGAATTTTTTTTTTTTTTTT 3']; double-stranded
cDNA was ligated to Hind III adaptors (Pharmacia), digested with
Not I and directionally cloned into the Not I and Hind III sites of
the Lafmid BA vector. Library went through one round of
normalization. Library constructed by Bento Soares and M.Fatima
Bonaldo.

ORGANISM Homo sapiens
EuXrayVote; Metazoa; Eumetazoa; Bilateria; Coelomata;

REFERENCE
AUTHORS

Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcophylli; Chonata; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Homidae; Homo.

TITLE
JOURNAL
COMMENT

The WashU-Merck EST Project
Unpublished (1995)

GDB: G00-415-907
Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watscn.wustl.edu
High quality sequence stops: 398
Source: IMAGE Consortium, LNLN
This clone is available royalty-free through LNLN; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Location/Qualifiers

FEATURES

1. 437
/organism="Homo sapiens"
/clone="43366"

BASE COUNT
ORIGIN

126 a 97 c 116 g 95 t 3 others

Query Match

Best Local Similarity 98.1%; Pred. No. 0.00e+00;
Matches 212; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

Db

220 TACCCCTGCCCTGAAAAAAGTGGCCAGCGCTGCTTACCCAGATCCCTCAAGAGAACCC 279

QY

1 TACCCCTGCCCTGAAAAAAGTGGCCAGCGCTGCTTACCCAGATCCCTCAAGAGAACCC 60

Db

280 AATGGCCAAAGGCTTGGCCAGATTCAGAACCAAGAGAGAGTATCCCATCCCGGGCTGGA 339

QY

61 AATGGCCAAAGGCTTGGCCAGATTCAGAACCAAGAGAGAGTATCCCATCCCGGGCTGGA 119

Db

340 TATCCCTGTGGGAGANATCATCTGTGAGAGAACCAAGAGAGAGTATCCCATCCCGGGCTGGA 399

QY

120 TATCCCTGTGGGAGANATCATCTGTGAGAGAACCAAGAGAGAGTATCCCATCCCGGGCTGGA 178

Db

400 TAGAGAGATTTGACGTGGGAGAGTGAACACGGA 435

QY

179 TAGAGAGATTTGACGTGGGAGAGTGAACACGGA 214

RESULT

11 AA327316 248 bp mRNA EST 20-APR-1997

LOCUS

EST10628 Colon I Homo sapiens cDNA 5' end.

DEFINITION

AA327316

ACCESSION

NID 91979561

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 248)

AUTHORS

Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
Bull,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man,Wai,C.,
Clifton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghegan,N.S.,
Glodek,A., Gnehm,C.L., Hanna,M.C., Healdrom,E., Hinkle,P.S., Jr.,
Kelley,J.M., Kelley,J.C., Liu,L.-T., Marmaro,S.M., Merrick,J.M.,
Moreno-Palancues,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,D.L., Saudek,D.M., Shirley,R.,

TITLE
JOURNAL
MEDLINE
COMMENT

Small,R.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
Bedarick,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.R.,
Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H.,
Raymond,L., Wei,Y.F., Wang,J., Xu,C., Yu,G.L., Ruben,S.M.,
Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
Fraser,C.M. and Venter,J.C.
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl.), 3-174 (1995)

Other ESTs: THC181443
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
Seq primer: M13 Reverse.

FEATURES

Location/Qualifiers

BASE COUNT
ORIGIN

67 a 63 c 60 g 55 t 3 others

Query Match

Best Local Similarity 97.5%; Pred. No. 0.00e+00;
Matches 195; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db

1 CTCCTGTGAGCAGCGTGTGTAAGGCTATGAAGGCGCAACGACGATGAGAGCTCA 60

QY

398 CTCCTGTGAGCAGCGTGTGTAAGGCTATGAAGGCGCAACGACGATGAGAGCTCA 457

Db

61 AGCCCAAGAGAAAGTNTTCAGAGAGTTCAGAGCTGACTTCAAAATTTCTGAGAGTCA 120

QY

458 AGCCCAAGAGAAAGTNTTCAGAGAGTTCAGAGCTGACTTCAAAATTTCTGAGAGTCA 517

Db

121 TCGCTAGTGGAGCAACCACTTCATGACCAAGCTGGCTCCATTCCTGTAATGCG 180

QY

518 TCGCTAGTGGAGCAACCACTTCATGACCAAGCTGGCTCCATTCCTGTAATGCG 577

Db

181 TGAAGGGGGAACATTAGC 200

QY

578 TGAAGGGGGAACATTAGC 597

RESULT

12 AA230715 503 bp mRNA EST 26-FEB-1997

LOCUS

MM04908.r1 Soares mouse 3NME12 5 Mus musculus cDNA clone 669758 5'

DEFINITION

similar to TR.G1184699 G1184699 TYROSYL-TRNA SYNTHETASE. ;

ACCESSION

AA230715

KEYWORDS

EST.

SOURCE

house mouse.

ORGANISM

Mus musculus

REFERENCE

1 (bases 1 to 503)

AUTHORS

Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Giesel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,

Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
The WashU-HMI Mouse EST Project
Unpublished (1996)

TITLE
JOURNAL
COMMENT

Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LNLN; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:409462
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 434.
Location/Qualifiers
1. .503
/organism="Mus musculus"
/strain="C57BL/6J"
/note="vector: p773D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTCACATCTGAAAGTGGAGCGCGCCCTATTTTCTTTT 3'],
on total mouse RNA [provided by Minoru Ko, Wayne State
Univ.]; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified p773 vector.
Library went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bernaldo."
/db_xref="taxon:10090"
/clone="669758"
/clone_lib="Soares mouse 3MME12 5"
/sex="unknown"
/dev_stage="12.5dpc total fetus"
/lab_host="DH10B"
1. .>503

FEATURES
SOURCE

Query Match 27.6%; Score 165; DB 17; Length 503;
Best Local Similarity 89.4%; Pred. No. 2,646-283;
Matches 193; Conservative 0; Mismatches 22; Indels 1; Gaps 1;

Db 288 TATCCAGCCCTGTAAGAGCTGGCCAGTGCCTACCACTCTTAAAGCAGAACCC 347
|||||
QY 1 TACCCCTGCTGTAAGAAATGCGCAGCGCTGCTACCACTCTTAAAGCAGAACCC 60
|||||

Db 348 TCTGCAAGAGCCCTGCAAGACTGCAAGCAGAGAGAGTCCATCCCGGCTGGA 407
|||||
QY 61 AATGGCCAAAGGCC-TGCCAAGAAATTCGAAGCAGAGAGTCCATCCCGGCTGGA 119
|||||

Db 408 TATCCGTGTGGGCAAAATCTCAGCGTGGAGAGCACCAGATGAGATAGCTTTATGT 467
|||||
QY 120 TATCCGTGTGGGAAATCATCATCTGTGAGAGAGCACCAGATGAGACCGCTGTATGT 179
|||||

Db 468 GGAGAAGATTATGTGGGGGAGAGCTGAACCCAGAGAC 503
|||||
QY 180 AGAGAAGATTGACGTGGGGAGAGCTGAACCCAGAGAC 215
|||||

LOCUS 13 H22113 209 bp mRNA EST 06-JUL-1995
DEFINITION Y134906.r1 Homo sapiens cDNA clone 160186 5'.
ACCESSION H22113
NID 9890808
KEYWORDS EST.
SOURCE human clone-160186 library-Soares breast 3NBHST vector-p773D
(Pharmacia) with a modified polylinker host-DH10B (ampicillin
resistant) primer-M13RP1 Rsite1-Not I Rsite2-Eco RI Adult human.
1st strand cDNA was primed with a Not I - oligo(dT) primer [5'

TGTTCACATCTGAAAGTGGAGCGCGCCCTATTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia),
digested with Not I and cloned into the Not I and Eco RI sites of a
modified p773 vector (Pharmacia). Library went through one round
of normalization to a Cot = 20. Library constructed by Bento Soares
and M. Fatima Bernaldo.
Homo sapiens
ORGANISM

REFERENCE
AUTHORS

Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
High quality sequence stops: 162
Source: IMAGE Consortium, LNLN
This clone is available royalty-free through LNLN; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Location/Qualifiers
1. .209
/organism="Homo sapiens"
/clone="160186"

FEATURES
SOURCE

Query Match 27.3%; Score 163; DB 11; Length 209;
Best Local Similarity 97.7%; Pred. No. 5,016-279;
Matches 169; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Db 1 GCTATGAAAGGGCCACACAGATGAGAGCTCAAGCCCAAGAGAGTCTTCAGAGT 60
|||||
QY 425 GCTATGAAAGGGCCACACAGATGAGAGCTCAAGCCCAAGAGAGTCTTCAGAGT 484
|||||

Db 61 TGACGCTGACTTCAAAATTTCTGAGAGTGCATCGC-CAGTNGANCACCAACTTCA 119
|||||
QY 485 TGACGCTGACTTCAAAATTTCTGAGAGTGCATCGCAGACAGTGAAGCAACCAACTTCA 544
|||||

Db 120 TGACAGCTGGGCTCCATTTCTGTAATTCGCTGAAGGGGGGGAACATTAGC 172
|||||
QY 545 TGACCAAGCTGGGCTCCATTTCTGTAATTCGCTGAAGGGGGGGAACATTAGC 597
|||||

LOCUS 14 AA049626 500 bp mRNA EST 09-SEP-1996
DEFINITION m36006.r1 Soares mouse embryo NMME13.5 14.5 Mus musculus cDNA
ACCESSION AA049626
NID 91529297
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Mus.

REFERENCE
AUTHORS

1 (bases 1 to 500)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.

(TM)

Distribution rights by Oxford Molecular Ltd

.....NEMTKLGSI~~SC~~SLKGNIS 168

Gap 11

195121 seqs, 42852602 residues

Listing first 45 summaries

```
1:pir1 2:pir2 3:pir3 4:pir4
```

Mean 36.208; Variance 118.180; scale 0.306

and is derived by analysis of the total score distribution.

SUMMARIES

This is a DE line. 3.12e+00

This is a DE line. 2.96e+01

ALIGNMENTS

ID	A55053	STANDARD;	PRT;	310 AA
----	--------	-----------	------	--------

AC XXXXXXXX
xy

DT 01-JAN-1900
YY

DE
XX

CC A: Molecule type: pr
CC a, accession: A44032

A:Experimental source: methylcholanth

SEQUENCE 310 AA; 33997 MW; 487802 CN;

Query Match 49.78; Score 596;

Matches	86;	Conservative	33;	Mismatches	35;	Indels	5;	Gaps	4
---------	-----	--------------	-----	------------	-----	--------	----	------	---

Db 148 IDASRLDRIGCIVTAKKHPDADSLVEEVDVGEAPRTVSGLVNHVPLEQMQRNVVL 207

QY 3 VIPSRDLIRVGKLIITVEKHPDADSLYEKIDVGGEAEPRTVVSGLVQFVPKEELQDRLVV 62

Db 208 LCNLKPAKMKGVLSQLAMVMCAS--SPEK-VEILAPNGSVPGDRITFDAPP-GEPPKEIN 2633

63 LCNLKPKMGRVESQGMILCASI EGINRQVEPLDPAGSAPGEHVFVKGYEKGPDEELK 122

DB 264 PKKKIWEIQPDLHINAECVAIYKGAPEVK-GKGVCR 301

QY 123 PAKKVE ENLQADEKISSECLAQMAQINEMIKLGSI CSNS 101

RESULT 2

[illegible]

.....

XX

XX

CC A; Reference number: A55053

CC A;Status: preliminary

CC A;Residues: 1-310 <KAO>
CC A;Cross-references: GB:U10117
SQ SEQUENCE 310 AA; 34253 MW; 494210 CN;

Query Match	49.18;	Score 588;	DB 2;	Length 310;
Best Local Similarity	53.88;	Pred. No. 1.54e-57;		
Matches	84;	Conservative	32;	Mismatches 35; Indels 5; Gaps 4

```

Db      151 SRDLRIGCITARKHPDADSLVEEDVGEIARPTVSSLVNHPLEOMONRMVILCN 210
      ||||:|: ||| ||||| ||||| ||||| ||| ||||| ||| ||||| |||
QY      6  SRDLRIGCITARKHPDADSLVEEDVGEIARPTVSSLVNHPLEOMONRMVILCN 210

```

```

D5      211 LKPAKMRGVLSQAMVWCAS--SPEK-ILILAPNGSVPGGRITFAFP-GEPPDKELNPKK 266
      ||| ||||| ||::||| : : : ||| ||::: : : ||| ||::|
Q7      66 LKPGKMRGVESQGMILCASITGGINROYEPLDPPAGSAPGEHVFVAKYGEKGPDEELPKK 125

```

DB 26/ KIWEDIQPDLHTNDECVATYKGVFEVK-GKVCRA 301
| : | : | : | : | : | :
QY 126 KVFEQLQADFKISEECIAQWKQTINFMTKLGSISCS 161

RESULT	3		
ID	S64113	STANDARD;	PRT; 376 AA
XX			

01-JAN-1900
This is a DE line.

TOIG of: s64113 check: 7848 from: 1 to: 376

CC >P1:564113
CC G4P1 protein - yeast (*Saccharomyces cerevisiae*)
CC N:Alternate names: protein G3085; protein YGL105w
C/Species: *Saccharomyces cerevisiae*

```
CC C/Date: 17-May-1996 #sequence_revision 17-May-1996 #text_change 12-Dec-1997
CC C/Accession: S64113; S69425
CC R/Castagnoli, L.; Paoluzzi, S.; Minenkova, O.
CC submitted to the Protein Sequence Database. May 1996
```

```
CC A:Reference number: S64112
CC A:Accession: S64113
CC A:Molecule type: DNA
CC A:Residues: 1-376 <CAS>
```

Cross-references: EMBL:U06971; NID:G13JZ2647; PID:e243893;
Experimental source: strain S288C
R.Castagnoli, L.; Paoluizi, S.; Minenkova, O.
submitted to the EMBL Data Library, April 1996

MIPS:YGI32648; MIPS:YGI32648; YGI32648

CC A/Accession: S69425
CC A/Molecule type: DNA
CC A/Residues: 1-376 <CAM>

CC C:genetics: MIM:607047; NID:91310/10/

CC A:gene: SGD:ARCL1; G4P1

CC A:cross-references: MIPS:YGL105w; SGD:S0003073

Query Match	38.6%	Score 462;	DB 2;	length 376,
Best local similarity	30.0%			

Matches	68;	Conservative	26;	Mismatches	38;	Indels	4;	Gaps	3;
Db	206	PSAIDFRVGIQAIKHPPADSLYSTIDVGDEGPRIVCSGLYKHFPIDAMQERYVVV	265						

QY 5 PSRLDIRVGKIITVEKHPDADSLYEKIDVGEAE-PRTVSGLVQFVPKKEELQDLRLVYVL 63
Db 266 CNLKPVNRGKISTAMVLCGSND--DK-VEFVEPPKDSKAGDKVFEFEGFGDEAPMKQLNP 3722

QY 64 CNLKPRKMGVESGMLLCASIEGINRQVEPLDPAGSAGEHVFVKGYEKQPDDELKP 123
Db 323 KKKIWEHLQPHFTTND 338
||| . ||| . . .

QY 124 KKKVEKLQADEFKISE 139

RESULT	4		
ID	G65095	STANDARD;	PRT; 110 AA.
XY			

DT 01-JAN-1900
XX This is a DE line.
DE
YY

CC A:Title: The complete genome sequence of *Escherichia coli* K-12
CC
CC A:Reference number: A64720. MIMD-97426617
CC

```
CC A/Accession: G65035
CC A/Status: preliminary; nucleic acid sequence not shown; translation not shown
CC A/Molecule type: DNA
CC A/Residues: 1-110 <RIAM>
```

CC A:Cross-references: GB:AEO00389; GB:U00096; NID:g1789451; PID:g1789445; UWGP:DB3
CC A:Experimental source: strain K-12, substrain MG1655
CC C:Genetics:
CC A:Gene: *YqjH*
CC

Query Match	25.78;	Score 308;	DB 2;	Length 110;
Best Local Similarity	50.68;	Pred. No. 4,30e-23;		
Matches 41;	Conservative 15;	Mismatches 24;	Indels 1;	Gaps 1;

D6
D7
D8
D9
D10 ARLEMRVGKIVEKKRHNADKLIVQVDVQOKTLQVTSLVPYYSSEELMGKITVVLCN 68
:::|||||: | : |||||: |||||: ||| | : : |||||
6 SRDLIRVGKIITVEKHDPADSLYEKIDVGEAPFRVYSGLVQFVPKEELQDRLLVVCN 65

```

Db      69 LQAKMRGETSECMLCAETD 89
      1 11111:11111:
QY      66 LKPKMRGVESQGMLLCASIE 86

```

RESULT	5	STANDARD;	PRT;	616 AA.
ID	SYTMMT			

AC	xxxxxxx
XX	
DT	01-JAN-1900
XX	

```
DE This is a DE line.
XX
CC A:Accession: A39517
CC A:Molecule type: DNM
```

CC A;Residues: 1-616 <NUR>
CC A;Cross-references: GB:M64273; GB:J05744; NID:g155135; PID:g155136
CC A;Experimental source: strain HB8, ATCC 27634
CC A;Note: the authors translated the codon 232 as th-

CC
C:Genetics:
CC
A:Gene: mets
CC
C:Superfamily: methionine-tRNA ligase
C:Keywords: aminoacyl-tRNA synthetase
AMP, homodimer, ligase, protein, homometh

Query Match	22.2%;	Score 266;	DB 1;	length 616;
Best Local Similarity	47.6%;	Pred. No. 3,87e-18;		
sq	SEQUENCE	616 AA;	70638 MW;	1863016 CN;

Accession	Matches	Conservative	Mismatches	Indels	Gaps
520 AKVELRYAEVLAEKHPNADRLVLRLSGNEE-RTVYSGIAKMYRPEELVGKKVVLVAN 578	39	22	19	2	

Qy 0 SKLDIKVGNLIVERKHPDADSLYEK
Db 579 LKPAKLRGIESQGMIL-AAQEG 599
||| |:|:|:|:|:|:| |:|

RESULT	6	
ID	D69431	STANDARD; PRT; 658 AA

XX AC xxxxxx
XX DT 01-JAN-1900
XX DE This is a DE line.
XX
CC A:Accession: D69431
CC A:Status: preliminary; nucleic acid sequence not shown; translation not shown
CC A:Molecule type: DNA
CC A:Residues: 1-658 <KLF>
CC A:Cross-references: GB:AE000782; TIGR:AF1453
SQ SEQUENCE 658 AA; 75695 MW; 2215053 CN;

Query Match 22.0%; Score 263; DB 2; Length 658;
Best Local Similarity 44.1%; Pred. No. 8 67e-18;
Matches 45; Conservative 19; Mismatches 32; Indels 6; Gaps 5;

Db 561 KLDIRIGVLAKEAKYKSKRLIKLIIDIGD-DQRIVSGIADYTP-EBLEGRVYVLN 618
OY 7 RLDIRVGKIITVEKHPDADSLYVEKIDVGEAEPRIVSGLVQ-FVPKEELQDRLVVVLGN 65
Db 619 LKPAFKGVESGMLAEKDG-KAV-LITPEKEVEPSTRV 657
OY 66 LKPKRMGVESQGMILCASIEGINRQVEPLDPPAGSAPGEHV 107

RESULT 7
ID G69177 STANDARD; PRT; 651 AA.
XX AC xxxxxx
XX DT 01-JAN-1900
XX DE This is a DE line.
XX
CC A:Accession: G69177
CC A:Status: preliminary; nucleic acid sequence not shown; translation not shown
CC A:Molecule type: DNA
CC A:Residues: 1-651 <MTH>
CC A:Cross-references: GB:AE000666
CC A:Experimental source: Strain Delta H
CC C:Genetics:
CC A:Gene: MTH587
CC A:Start codon: TTG
CC C:Keywords: aminoacyl-tRNA synthetase; protein biosynthesis
SQ SEQUENCE 651 AA; 74631 MW; 2052441 CN;

Query Match 18.5%; Score 222; DB 2; Length 651;
Best Local Similarity 44.7%; Pred. No. 4.59e-13;
Matches 42; Conservative 20; Mismatches 27; Indels 5; Gaps 5;
Db 558 MDIRGVIRSAIRIGSDKLLIIDVGEREMQ-VYAGLAETKSP-EDLVERKITVYVNL 615
OY 8 LDIRVGKIITVEKHPDADSLYVEKIDVGEAEPRIVSGLVQ-FVPKEELQDRLVVVLGN 66
Db 616 KPAKLGVSSEGMLV-ATGESLNI-LDPGDAEYG 647
OY 67 KPKRMGVESQGMILCASIEGINRQVEPLDPPAG 100

RESULT 8
ID F64457 STANDARD; PRT; 651 AA.
XX AC xxxxxx
XX DT 01-JAN-1900
XX DE This is a DE line.
XX
CC A:Authors: Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Weese, C.R.; Ve
CC A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jan
CC A:Reference number: A64300; MUID:96337999

CC A:Accession: F64457
CC A:Status: preliminary; nucleic acid sequence not shown; translation not shown
CC A:Molecule type: DNA
CC A:Residues: 1-651 <BUJ>
CC A:Cross-references: GB:U67567; GB:L77117; NID:g1591897; PID:g1591898; TIGR:MJ12
CC C:Genetics:
CC A:Map position: FOR1205666-1207621
CC C:Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis
SQ SEQUENCE 651 AA; 75558 MW; 2233744 CN;

Query Match 17.6%; Score 211; DB 2; Length 651;
Best Local Similarity 40.0%; Pred. No. 8.06e-12;
Matches 30; Conservative 20; Mismatches 24; Indels 1; Gaps 1;

Db 553 KIDLVGVEVEEDIPKSKKLKLVNDIGD-EKRIVSGIGGYKPEPDLVCKRVIVIONL 611
OY 7 RLDIRVGKIITVEKHPDADSLYVEKIDVGEAEPRIVSGLVQFVPKEELQDRLVVVLGN 66
Db 612 KPAKLGVSSEGML 626
OY 67 KPKRMGVESQGMIL 81

RESULT 9
ID I64113 STANDARD; PRT; 682 AA.
XX AC xxxxxx
XX DT 01-JAN-1900
XX DE This is a DE line.
XX
CC A:Authors: Geophagen, N.S.M.; Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser,
CC A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae
CC A:Reference number: A64000; MUID:95350630
CC A:Accession: I64113
CC A:Status: nucleic acid sequence not shown; translation not shown
CC A:Molecule type: DNA
CC A:Residues: 1-682 <TIG>
CC A:Cross-references: GB:U32807; GB:LA2023; NID:g1574725; PID:g1574731; TIGR:HI12
CC C:Genetics:
CC A:Gene: metG
CC C:Superfamily: methionine-tRNA ligase
CC C:Keywords: aminoacyl-tRNA synthetase; ATP; ligase; protein biosynthesis
SQ SEQUENCE 682 AA; 77022 MW; 2278048 CN;

Query Match 16.7%; Score 200; DB 2; Length 682;
Best Local Similarity 38.1%; Pred. No. 1.38e-10;
Matches 32; Conservative 24; Mismatches 24; Indels 4; Gaps 4;
Db 582 AKLDRVAKVLCGEAVPESNKLREPLDGD-HTROYSGIKAAVNRPELEGRVINYA 640
OY 6 SRLDIRVGKIITVEKHPDADSLYVEKIDVGEAEPRIVSGLVQFVPKEELQDRLVVVLGN 64

Db 641 NLAPRKRFV-SEGMIISAGTGG 663
OY 65 NKKPKRM-GVESQGMILCASIEG 87

RESULT 10
ID I16682 STANDARD; PRT; 649 AA.
XX AC xxxxxx
XX DT 01-JAN-1900
XX DE This is a DE line.
XX
CC A:Accession: I16682
CC A:Status: preliminary
CC A:Molecule type: DNA
CC A:Residues: 1-649 <MEC>
CC A:Cross-references: EMBL:X57925; NID:g39988; PID:g39989

CC C:Genetics:
CC A:Gene: mets
CC C:Keywords: aminoacyl-tRNA synthetase; ATP; ligase; protein biosynthesis;
SEQUENCE 649 AA; 74354 MW; 2141119 CN;

Query Match	15.7%;	Score 188;	DB 2;	Length 649;
Best Local Similarity	37.8%;	Pred. No. 2.95e-09;		
Matches	31;	Conservative	23;	Mismatches 27; Indels 1; Gaps 1;

D_b 551 AKVDLRVAEYVQPERKMNADKLKLQDLDG-GEKRQVISGAFFYPEELIGKKVICVAN 609
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
Q_y 6 SRUDIRVGKIITVEKHDPDASLVEKIDVGEAEPRTVSGIVQFVKKEELDRLVWVLGN 65

```

Db      610 LKPAKLRGEWSEGMILTAGSGG 631
      ||| |::| |::|::|
QY      66 LKPQKRGVESQGMILCASIEG 87

```

RESULT	11	STANDARD;	PRT;	664 AA..
ID	S66067			
XY				

DT 01-JAN-1900
XX
DE This is a DE line.

CC A:Cross-references: EMBL:ID26185; NID:g467332; PID:d1005815; PID:g467427
CC A:Note: the nucleotide sequence was submitted to the EMBL Data Library,
CC Genebank:

CC C;keywords: ligase
SQ SEQUENCE 664 AA; 76187 MW; 2333297 CN;

Query Match	15.1%;	Score 181;	DB 2;	Length 664;
Best Local Similarity	39.5%;	Pred. No. 1.73e-08;		
Matches	34;	Conservative	18;	Mismatches 31; Indels 3; Gaps 3;

Dd 568 VELRYAEVIEAPVKKADRLKLQDLDG-FEKROYVSGIKHYTPBELVGKKLYCVT-NL 623
::||: | | | :|: | |||: :: | : || | |
Qy 8 LDIRGKIIVTEKHPRDADSLYEKIDVEAEPRYVSSGLV-QFVPRELDORLNVVLICNL 66

```
Db      626 KPVKLRGELSQGMILAGEADVLKV 651
        || |:|| |||:| : :| : |
QY      67 KPQMRGVESQGMLLCASTIEGINRV 92
```

RESULT	12		
ID	E69657	STANDARD;	PRT; 664 AA.

DT 01-JAN-1900
XX
DE This is a D
XX

CC	A/Experimental source: strain 168
CC	C/Genetics:
CC	A/Gene: mets

SO	SEQUENCE	664	AA;	76187	MM;	2333297	CN;	
	Query Match			15.1%		Score 181;	DB 2;	Length 664;

Matches 34; Conservative 18; Mismatches 31; Indels 3; Gaps 3;

Db 568 VELRVAEVTAEAPVKKADRLIKQLDLC-FEKRRQVSGIAKHTPEELVGKKLYCVT-NL 625

QY 8 LDINVGKIIIEKHPADSLYIEKIDVGEAEPRIVSGLV-QFVPEKEELODRLVVILCNL 66

Db 626 KPVKLRGELSGMILAGEADGVLKVV 651

67 KPOKMRGVESQGMILLCASIEGINRQV 92

RESULT	13
ID A64572	STANDARD; PRT; 650 AA

AC	XXXXXX
XX	
DT	01-JAN-1900

DE This is a DE line.
XX
CC A;Authors: Hayes,
CC
CC

CC	A:Reference number: A64520; MUID: 97394467
CC	A:Accession: A64572
CC	A:Status: preliminary; nucleic acid sequence not shown; translation not shown

```

CC A:Residues: 1-650 <TOM>
CC A:Cross-references: GB:AE000557; GB:AE000511; NID:g2313514; PID:g2313521; TIGR:
CC C:Keywords: ligase
CC

```

Query Match	14.4%;	Score 173;	DB 2;	Length 650;
Best Local Similarity	35.5%;	Pred. No. 1.29e-07;		
Matches	27;	Conservative	23;	Mismatches 24;
			Indels	2;
			Gaps	2

Dd 552 KVEIKVGLLKEAQRLEKSNKLIRLKVDLGEGRKQIISIALDIYEP-ESLVGOMCVVAN 610
:::| | : : | :| :| : :| :| :| :| :| :
Qy 7 RLDIRVGKITVEKHDPDASLYEKKIDVGEABRTVSGSL-VGFYPKELQDLRLVVVLGN 65

```

Db      611 LKPAKLMGEMSEGMIL 626
      ||| | : | | : || : |
QY      66 LKPKMRGVESQGMILL 81

```

RESULT	14	STANDARD;	PRT;	677	AA
ID	SYECMT				

AC	xxxxxx
XX	
DT	01-JAN-1900
XX	

DE This is a DE line.
XX
CC A:Contents: annotation; X-ray crystallography, 2.5 angstroms
C:Contents:

CC	A;Gene: metG
CC	A;Map position: 46 min
CC	C;Function:
CC	A;Description: catalyz

CC A;Pathway: protein biosynthesis
CC C;Superfamily: methionine--tRNA ligase
CC C;Keywords: aminocacyl-tRNA synthetase; ATP; homodimer; ligase; metalloprotein;
E;2-577;product: methionyl-tRNA ligase;+status: predicted;+unpub

```
CC F:62-66,137-157,335-341,434-438/Region: tRNA 3'-acceptor end binding
CC F:452-468/Region: anticodon recognition
CC F:16,22-55,336/Active site: Tyr, His, Asp, Lys #status predicted
CC F:146,149,159,162/Binding site: zinc (Cue) #status experimental
```

Query Match	13.4%	Score 160;	DB 1;	Length 677;
Best Local Similarity	33.7%	Pred. No. 3	19e-06.	
SQ	SEQUENCE	677 AA;	76254 MW;	2257999 CN;

Matches 35; Conservative 26; Mismatches 37; Indels 6; Gaps 6

Db 577 AKVDLRVALIENAEFVEGSDKILRLTLDLG-GEKNVFGSIRSAYPDPALIGRHTIMVA 635
.....

QY 6 SRLLDIRVGGKIITVEKHHPDADSLVYEKIDVGEAPRIVVSGL-VQFVPRKEQLQDRLLVVLIC 64

Db 636 NLAPPKMRFGI-SEGVMYAG-PG-GKDIIFLLSPDAGAKPGHOV 676

QY 65 NLKPQKMR-GVESQGMILCASIEGINRQVEPLDPYPAGSAPGEHV 107

[illegible]

Search completed: Sat Apr 18 02:05:28 1998
Job time : 54 secs.

THIS PAGE BLANK (USPTO)